

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 25, 2001, 11:25:36 ; Search time 82.56 Seconds
(without alignments)
8921.131 Million cell updates/sec

Title: US-09-435-054-1
Perfect score: 1173
Sequence: 1 ccacgctcgcaccacac.....tcaaaaaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0601:*

1: /SIDS1/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SIDS1/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SIDS1/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SIDS1/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SIDS1/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SIDS1/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SIDS1/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SIDS1/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SIDS1/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SIDS1/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SIDS1/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SIDS1/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SIDS1/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SIDS1/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SIDS1/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SIDS1/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SIDS1/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SIDS1/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SIDS1/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SIDS1/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1173	100.0	1173	21	Maize LEC1 # 1 cod
2	226.4	19.3	1098	21	Wheat LEC1 # 3 cod
3	200	17.1	1121	21	Soybean LEC1 # 2 c
4	195.8	16.7	796	21	Soybean LEC1 # 3 c
5	192.4	16.4	942	21	Soybean LEC1 # 1 c
6	179	15.3	3395	19	Arabidopsis leafy-3.4 kb BstVI fragm
7	179	15.3	3395	21	Arabidopsis leafy-7.4 kb BspRI fragm
8	179	15.3	7560	19	Arabidopsis leafy-cDNA sequence of a
9	179	15.3	7560	21	
10	177	15.1	627	19	
11	177	15.1	627	21	

12	175	14.9	905	21	Veronia mespilifol
13	166.2	14.2	622	21	Maize LEC1 # 3 cod
14	154.2	13.1	470	21	Pinus radiata tran
15	154.2	13.1	561	21	Pinus radiata tran
16	149.8	12.8	1014	21	Pinus radiata tran
17	148	12.6	481	21	Argemone mexicana
18	133.2	11.4	382	21	Eucalyptus grandis
19	128.6	11.0	376	21	Eucalyptus grandis
20	121.2	10.3	372	21	Pinus radiata tran
21	120	10.2	739	21	Pinus radiata tran
22	109.8	9.4	815	21	Pinus radiata tran
23	108.6	9.3	768	21	Fusarium venenatum
24	107	9.1	433	21	Arabidopsis thalia
25	100.4	8.6	1284	12	Eucalyptus grandis
26	100.2	8.5	407	21	Pinus radiata tran
27	99.4	8.5	417	21	Pinus radiata tran
28	99.4	8.5	758	21	Arabidopsis thalia
29	87.4	7.5	238	21	Arabidopsis thalia
30	83.6	7.1	763	21	Arabidopsis thalia
31	82.4	7.0	335	21	Maize LEC1 # 2 cod
32	78.8	6.7	445	21	Pinus radiata tran
33	73.4	6.3	1900	21	Pinus radiata tran
34	66.4	5.7	114955	20	Human adenosine Al
35	65.6	5.6	1635	13	Mutant thermostabl
36	65.6	5.6	1899	13	Mutant thermostabl
37	65.6	5.6	2043	13	Mutant thermostabl
38	65.6	5.6	2277	13	Mutant thermostabl
39	65.6	5.6	2367	13	Mutant thermostabl
40	65.6	5.6	2493	13	Mutant thermostabl
41	65.6	5.6	2512	16	Tfif DNA-polymerase
42	65.6	5.6	4947	16	Tsps17 polymerase
43	63.6	5.4	1089	18	Elmeria tenella sp
44	60	5.1	114955	20	Human adenosine Al
45	59.4	5.1	58857	21	Nucleotide sequenc

ALIGNMENTS

RESULT 1
AAAA27450

ID AAA27450 standard; cDNA; 1173 BP.

XX

AC AAA27450;

XX

DT 11-SEP-2000 (first entry)

XX

DE Maize LEC1 # 1 coding sequence.

XX

KW Maize: leafy cotyledon 1 transcriptional activator; LEC1; apomixis;

KW selectable marker; transgenic plant; transgenic seed; HAP3; ss.

XX

OS Zea mays.

XX

PH Key

FT CDS

FT

FT

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Location/Qualifiers
69...905
/*tag= a
/product= "Maize LEC1"

WO200028058-A2.

18-MAY-2000.

09-NOV-1999; 99WO-US26514.

09-NOV-1998; 98US-0107643.

10-NOV-1998; 98US-0107810.

(PION-) PIONEER HI-BRED INT INC.

(DUPO) DU PONT DE NEMOURS & CO E I.

Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;

Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;

CC CCAAT-box binding protein. LEC1 expression initiates the formation of
 CC embryo-like structures and improves growth and recovery of transformants.
 CC When apomixis occurs, i.e. the replacement of sexual reproduction by
 CC asexual reproduction, LEC1 expression in the nucellus integument, or
 CC cell specific expression in the megaspore mother cell would trigger
 CC embryo formation from maternal tissues only. This results in the
 CC production of seeds identical to the parent. Using LEC1, transgenic high
 CC yielding seeds could be developed. In addition, LEC1 could be used for
 CC positive selection of a transformed cell (transgenic plant), for
 CC increasing transformation efficiency and for increasing recovery of
 CC regenerated plants.
 XX
 SQ Sequence 796 BP; 208 A; 174 C; 188 G; 226 T; 0 other;,
 Query Match 16.7%; Score 195.8; DB 21; Length 796;
 Best Local Similarity 75.9%; Pred. No. 3.2e-28;
 Matches 242; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
 QY 159 gcggcgccggatccgcgagcagcagcggcgtgatccgcatcgcgaacgtgacccatc 218
 Db 7 gcaatggcggagtgagggaacagaccagatcatatccgatagcgaacgtgataagatc 66
 QY 219 atcgccgcgtgtgtccgcgcgcgcgaagatctcggacacgcccagagagacgacccag 278
 Db 67 atgcgtcgattctgccagcgcgcgcgaagatctcagacacgagcagcagagagacccag 126
 QY 279 gactgctgcgagatcatcagcttcacgcgggagcgcgaacacgagcgtgcccagcgg 338
 Db 127 gactgctgcgagatcatcagcttcacgcgggagcgcgaacacgagcgtgcccagcgg 186
 QY 339 gacgagcgcgaagaccatccacgcgcgcgcgaagcgtgtgtggccatgagccgcctgcctc 398
 Db 187 gacgagcgcgaagacggtgacgcgcgcgcgaagcgtgtgtggccatgagccgcctgcctc 246
 QY 399 gacgactacgctgc 458
 Db 247 gacgactacgctgc 458
 QY 459 gc 477
 Db 307 cctgcttctgcagacgc 325
 RESULT 5
 AAA27456
 ID AAA27456 standard; cDNA: 942 BP.
 XX AC AAA27456;
 XX 11-SEP-2000 (first entry)
 DT DE Soybean LEC1 # 1 coding sequence.
 XX KW Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3; ss.
 XX OS Glycine max.
 XX FH Key Location/Qualifiers
 FT CDS 3..725
 FT /*tag= a
 FT /product= "soybean LEC1 # 1"
 FT /partial
 XX WO200028058-A2.
 XX 18-MAY-2000.
 XX 09-NOV-1999; 99WO-US26514.
 XX 09-NOV-1998; 98US-0107643.
 XX 10-NOV-1998; 98US-0107810.
 XX (PION-) PIONEER HI-BRED INT INC.
 XX (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX WPI: 2000-376568/32.
 DR P-PSDB: AAY96221.
 XX New HAP3-type CCAAT-box binding transcriptional activators,
 XX particularly Leafy cotyledon 1 transcriptional activator, useful for
 XX inducing somatic embryogenesis or apomixis in a plant cell -
 XX Claim 1; Page 90-91; 94pp; English.
 XX The present sequence is the coding sequence of soybean leafy cotyledon 1
 XX transcriptional activator, LEC1. This sequence encodes a HAP3-type

QY 131 caacaatgctgcgtctcagcagcagcagc 190
 Db 131 cagcaaccacagtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 190
 QY 191 gatcgcatcgcgaacgtgatccgcacatcgc 250
 Db 191 catgccaatcgtcaacgtgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 250
 QY 251 ctgcgacgacgccaagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 310
 Db 251 ccccgatgatgcaaggagagagacaaatcccaagagtgctgctgagtgatcatcagcttcacac 310
 QY 311 ggggagggccaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 370
 Db 311 cggggagggccaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 370
 QY 371 gctgtggccatgacgc 430
 Db 371 gctttgggcaatgagtaagcttgattcgacgactacatcagacccgttaaccatgtacct 430
 QY 431 ccacgcctaccgcgagttcgaggcgac 458
 Db 431 tcacgcctaccgtgactgagggtgac 458
 RESULT 4
 AAA27461
 ID AAA27461 standard; cDNA: 796 BP.
 XX AC AAA27461;
 XX 11-SEP-2000 (first entry)
 DT DE Soybean LEC1 # 3 coding sequence.
 XX KW Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3; ss.
 XX OS Glycine max.
 XX FH Key Location/Qualifiers
 FT CDS 1..516
 FT /*tag= a
 FT /product= "Soybean LEC1 # 3"
 FT /partial
 XX WO200028058-A2.
 XX 18-MAY-2000.
 XX 09-NOV-1999; 99WO-US26514.
 XX 09-NOV-1998; 98US-0107643.
 XX 10-NOV-1998; 98US-0107810.
 XX (PION-) PIONEER HI-BRED INT INC.
 XX (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX WPI: 2000-376568/32.
 DR P-PSDB: AAY96221.
 XX New HAP3-type CCAAT-box binding transcriptional activators,
 XX particularly Leafy cotyledon 1 transcriptional activator, useful for
 XX inducing somatic embryogenesis or apomixis in a plant cell -
 XX Claim 1; Page 90-91; 94pp; English.
 XX The present sequence is the coding sequence of soybean leafy cotyledon 1
 XX transcriptional activator, LEC1. This sequence encodes a HAP3-type

29-DEC-1999. 24-JUN-1999; 99WO-US14384.
 24-JUN-1998; 98US-0103478.
 17-NOV-1998; 98US-0193931.
 (REGC) UNIV CALIFORNIA.
 Harada JJ, Lotan T, Ohto M, Goldberg RB, Fischer RL;
 WPI; 2000-150588/14.
 New embryo-specific gene useful for producing transgenic plant
 Claim 19; Page 57-64; 69pp; English

The present sequence represents a 74 kb EcoRI fragment of a subcloned

LEC1 genes are thought to play a central role in late embryogenesis, in specifying cotyledon identity during embryo development. LEC1 polypeptides may act as transcription factors. LEC1 polynucleotides are

CC	development of embryonic tissue in a plant. In both cases, the LEC1
CC	policynucleotide is introduced into the plant through a sexual cross and
CC	is co-expressed in an antisense orientation with a second heterologous
CC	polynucleotide selected from AP2 and RAP2 genes of Arabidopsis. The
CC	LEC1 polynucleotide is also useful for targeting expression in a seed,
CC	and for preparing expression cassettes for suppressing or enhancing
CC	endogenous LEC1 gene expression, which is useful in weed control or
CC	for improving nutritional value of plant tissue respectively. LEC1
CC	polypeptides and polynucleotides are especially used for increasing or
CC	decreasing storage protein content in cotyledons or leaves. LEC1 also
CC	increases reproductive tissue mass, e.g., increases fruit size, seed
CC	mass, its protein or its oil.
XX	
SQ	Sequence 7560 BP; 2378 A; 1326 C; 1264 G; 2573 T; 19 other;
Query Match	15.3%; Score 179; DB 21; Length 7560;
Best Local Similarity	66.4%; Prid. No. 4.7e-25;
Matches 257; Conservative	0; Mismatches 130; Indels 0; Gaps
QY	71 ggaactccagcagcttccctcgccgcggcgaggagaatgctcggcggcgaggcgacg 130
Dd	4406 gcacagcaacaaccaccacccatagcacagctcagtcagttagccgcgcggtgacaa 4465
QY	131 caacaatggcgcgctgtctcagcagcatcgcgcgccgcggcagtcgccgagcagcgacct 190
Dd	4466 gaacaatggtatcgttgccaagcagaacacacattgtgtgctctgtagcaagaccataa 4525

4526 catgccaatcgcaaacgctcataagaatcatgcgtaaaaccttaccgtctcacgccaataat 4585

13C

z51 ctcggaacgcccagggaagacgagtcgagtcgaggtacatcagcttcacac 310

4586 ctctgacgacccaaagaacgaattcaagaattatctccgaatacatcagcttcgtgac 1645

311 ggagggccaacagcgtgccagcggaagaccatcacgcggacgt 370
|| ||||||| |||| | |||||| | ||||
4646 cagtqaaccaacagcttgcacaqtgacaactaaactctaagaatat 4705

[illegible][illegible]

4700 caaccggaacccgcagacagagaccga **4792**

CC decreasing storage protein content in cotyledons or leaves. LEC1 also
CC increases reproductive tissue mass, e.g., increases fruit size, seed
CC mass, its protein or its oil.

SQ Sequence 627 BP; 162 A; 153 C; 163 G; 149 T; 0 other;

Query Match	15.1%	Score 177;	DB 21;	Length 627;
Best Local Similarity	68.1%	Pred. No. 9.9e-25;		
Matches 246;	Conservative	0;	Mismatches 115;	Indels

QY	97	ccggcgcgagaatggtcgtcgcgcgggcgcccaaatggcgcgctgctcagcagc	156
Db	5	ccagctcagtcattagtagccggcgcggtgacaagaacaattggtatcggtgccagcagc	64
QY	157	atggggcgcgcgatccgcagacgagaccgctatccgatatcggaacgtagtcgcga	216
Db	65	aaccaccattgtgtggtctgtgagcaagaccatactccaatcgcaacgctcatagaa	124
QY	217	tcattgcggcggtgtcggcgcgacgccaaatctcgacagcgccaaagagacgacac	276
Db	125	tcattgcgtaaaacctaccgtctcacgccaaaattctgacgagcccaagaacgatttc	184
QY	277	aggadtgcgtgtcgagttacatcagcttcatcacgggggagcccaacgagcgtgcagc	336
Db	185	aagaatgtgtctcgagtacatcagcttcgtgaccggtgaagcccaacgagcgtgccac	244
QY	337	gggaacagcgcaagacatcacgcggaagcgtgctgtggccatgagcgcgctcggt	396
Db	245	gtgacaacgttaagaccataactgctgaagatactcttgggttatgagaacttgggt	304
QY	397	tcgagactactgtgagccgctcgcgctactctccacgctaccgcgagttcgaggcg	456
Db	305	tcgataactaactgtgaccctcaccgtgttctattaacccgttcctgtgatatgagaccg	364
QY	457	a	457
Db	365	a	365

RESULT 12

AAA27457
ID AAA27457 standard; cDNA; 905 BP.

AAA27457;

DT 11-SEP-2000 (first entry)

Veronia mespilifolia LEC1 coding sequence.

KW Leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
KW selectable marker; transgenic plant; transgenic seed; HAP3; ss.

OS *Veronia mespilifolia*.

Key	Location/Qualifiers
CDS	58..702
FT	/*tag= a
FT	/product= "LEC1"

PN WO200028058-A2.

18-MAY-2000.

09-NOV-1999; 99WO-US26514.

PR 09-NOV-1998; 98US-0107643.

XX
XX
1018/010-5006; 1966T-ADN AT

PA (PION-) PIONEER HI-BRED INT. INC.
PA (DUPO) DU PONT DE NEMOURS & CO. E. I.

XX
PI
Lowe KS, Gordon-Kamm WJ. Klein TM.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

PI XX DR DR XX

Sun X, Hoerster GJ, Gregory CA, Nadimpalli R; WPI; 2000-376568/32. P-PSDB; AAY96217.

New HAP3-type CCAAT-box binding transcriptional activators

new and specific type of plant cell wall binding transcriptional activators, particularly Leafy cotyledon 1 transcriptional activator, useful for inducing somatic embryogenesis or apomixis in a plant cell.

Claim 1; Page 82-83; 94pp; English.

The present sequence is the *Veronica mespilifolia* coding sequence of leafy cotyledon 1 transcriptional activator, LEC1. This sequence encodes a HAP3-type CCAAT-box binding protein. LEC1 expression initiates the formation of embryo-like structures and improves growth and recovery of transformants. When apomixis occurs, i.e. the replacement of sexual reproduction by asexual reproduction, LEC1 expression in the nucellus integument, or cell specific expression in the megaspore mother cell would trigger embryo formation from maternal tissues only. This results in the production of seeds identical to the parent. Using LEC1, transgenic high yielding seeds could be developed. In addition, LEC1 could be used for positive selection of a transformed cell (transgenic plant), for increasing transformation efficiency and for increasing recovery of regenerated plants.

Sequence 905 BP; 243 A; 184 C; 231 G; 247 T; 0 other;

Query Match 14.9%; Score 175; DB 21; Length 905;
Best Local Similarity 74.6%; Pred. No. 2.4e-24;

Qy 172 tccgcagcaggaccggctgatgccgatccqqaacqtatccqcatcatcgcgcgcgtgc 231

Db 197 **ttcgagaacaggaccgcttc**atgccgatagcaaacqtatccqcatcatcgcgaagatcc 256

Qy 232 tgcgggcacgccaaagatctcggacgacgcccaaggagacgattccaqqagatcactatcqa 291

Db 257 ttctccacatgccaaagatctctgatgatgccaaagagacgatccaaagaatgtgtttcag 316

QY 292 agtacatcagcttcatcacgggggagggccaacgagcggtgccagcgggagcagcagca 351

Db 317 agtacattagctttgtcacaggcgaggcaaatgaccgctgccagcgtgagcaaaaggaaqa 376

QY 352 ccataccgccgaggacgtgctgtgggcccatgagccgcctcggcttcgacqactacqtca 411

Db 377 ccatacagctgaagatgtgctctgggctatgagcaaaactgggatttgatqattatatc 436

QY 412 agccgctcggcgctacctccacgctaccgcgagttcgagggcgacgcgcgcg 466

Db 437 agcccttgactgtgtatctccatcgctacagggaggtttgatggtggcgaacgtgg 491

RESULT 13

AAA27459
ID AAA27459 standard; cDNA; 622 BP.

AAA27459;

DT 11-SEP-2000 (first entry)

DE Maize LEC1 # 3 coding sequence.

Maize; leafy cotyledon 1 transcriptional activator; LECL; apomixis; selectable marker; transgenic plant; transgenic seed; HAP3; ss.

Zea mays.

Key	Location/Qualifiers
CDS	3..620

```

/product= "Maize LEC1 # 3"

```

F^T /transl_except= (pos:234..236,aa:Xaa)

The present invention relates to novel plant transcription factors from *Eucalyptus grandis* or *Pinus radiata*. The present sequence is the coding sequence for one such transcription factor. The transcription factor may be used to produce a plant having modified gene expression such as a woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or to modify the activity of a polypeptide in a plant. The transcription factors of the present invention are members from the following families of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic helix-loop-helix zipper, homeotic/homeodomain/homeobox/WADS, homeodomain zipper, LIM domain, AP2

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:21:45 ; Search time 9.89 Seconds
(without alignments)
962.894 Million cell updates/sec

Title: US-09-435-054-2

Perfect score: 1514

Sequence: 1 MDSSFLPAAGNCSAAGG.....ASHTPQSGGLEHPPHPPAYK 278

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	391	25.8	179	1	CBFA_MAIZE
2	325.5	21.5	209	1	CBFA_PETMA
3	324.5	21.4	207	1	CBFA_HUMAN
4	324.5	21.4	207	1	CBFA_MOUSE
5	322	21.3	151	1	CBFA_CHICK
6	295.5	19.5	144	1	HAP3_YEAST
7	274	18.1	116	1	HAP3_SCHPO
8	265	17.5	205	1	HAP3_KLULA
9	232.5	15.4	122	1	CBFA_XENLA
10	195	12.9	159	1	DR1_ARATH
11	179	11.8	369	1	MAF2_RAT
12	176.5	11.7	370	1	MAF2_MOUSE
13	171	11.3	590	1	HMF2_DROME
14	170	11.2	369	1	TMAP_AVT54
15	161	10.6	176	1	TMAP_HUMAN
16	149	9.8	428	1	FXB2_MOUSE
17	147	9.7	1185	1	DRPL_HUMAN
18	139	9.2	421	1	BR3A_MOUSE
19	139	9.2	423	1	BR3A_HUMAN
20	135.5	8.9	1183	1	DRPL_RAT
21	134	8.9	549	1	DSX_DROME
22	132.5	8.8	497	1	FXD2_HUMAN
23	132.5	8.8	512	1	ANX7_XENLA
24	129.5	8.6	200	1	GRP_HORVU
25	129	8.5	440	1	FXGA_CHICK
26	129	8.5	553	1	FAC1_MOUSE
27	127.5	8.4	911	1	ANDR_MOUSE
28	124.5	8.2	895	1	ANDR_PANTR
29	124.5	8.2	919	1	ANDR_MACFA
30	123.5	8.2	269	1	MRG1_MOUSE
31	123.5	8.2	465	1	FXDL_HUMAN
32	123.5	8.2	895	1	ANDR_PAPHA
33	123	8.1	402	1	POXM_DROME

ALIGNMENTS

RESULT 1
CBFA_MAIZE

AC P25209; STANDARD; PRT; 179 AA.

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DE 01-OCT-1996 (Rel. 34, Last annotation update)

DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN

CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).

GN NFY2.

OS Zea mays (Maize).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae;

OC Andropogoneae; Zea.

OX NCBI_TaxID=4577;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-92195809; PubMed-1549471;

RA Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I.,

RA Benoist C., Mathis D.;

RT "Evolutionary variation of the CCAAT-binding transcription factor

NF-Y";

RL Nucleic Acids Res. 20:1087-1091(1992)

CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY

CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR

CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.

CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO

CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A

CC SUBUNIT, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN

CC DOMINANT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.

CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

CC -----

DR EMBL; X59714; CAA42234.1; --

DR HSSP; P19267; IBFM.

DR MaizeDB; 69282; --

DR InterPro; IPR000947; --

DR Pfam; PF00808; CBF2_NFYB_HMF; 1.

DR PRINTS; PR00615; CCAATSUBUNTA.

DR PROSITE; PS00685; CBF_A_NFYB; 1.

KW Transcription regulation; DNA-binding; Activator; Nuclear protein.

FT DOMAIN 1 29 A DOMAIN.

FT DOMAIN 30 119 B DOMAIN.

FT DOMAIN 120 179 C DOMAIN.

FT DNA_BIND 36 42 BY SIMILARITY.

FT SEQUENCE 179 AA; 18995 MW; 30621316CE469454 CRC64;

Query Match 25.8%; Score 391; DB 1; Length 179;
 Best Local Similarity 44.6%; Pred. No. 3.4e-19;
 Matches 82; Conservative 31; Mismatches 45; Indels 26; Gaps 5;

QY 8 PAAGA---ENGSAAGGANNAGAAQAAPAIREDRLMPIANVIRIMRVLPAHAHAKISDD 64
 DB 8 PGGGGSHESGPRGGGGG-----SVREQDRLPIANISRIKMKKAIPANGKIAKD 58
 QY 65 AKETQECVSEYISFITEANERQOREKTTITADVLWAMSLRGDDYVEPLGAYLHRY 124
 DB 59 AKETQECVSEYISFITEANERQOREKTTITADVLWAMSLRGDDYVEPLGAYLHRY 118
 QY 125 REFEGDARGVGLPGNAPRGDD-----HUPHSMSPAAAMLKSRGPVS-GAAMLPHHHH 176
 DB 119 REMEGDSK-----LTAKSSDGSIIKDALGHVGAASSAAEGMGQCGAYNQGMYQPOYH 172
 QY 177 HHDM 180
 DB 173 NGDI 176

RESULT 2
 CBFA_PETMA
 ID CBFA_PETMA STANDARD; PRT; 209 AA.
 AC P25210;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN CHAIN B)
 DE CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).
 OS Petromyzon marinus (Sea lamprey).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
 OC Petromyzoniformes; Petromyzontidae; Petromyzon.
 OX NCBI_TaxID=7757;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92195809; PubMed=1549471;
 RA Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I.,
 RA Benoist C., Mathis D.;
 RT "Evolutionary variation of the CCAAT-binding transcription factor
 NF-Y.";
 RL Nucleic Acids Res. 20:1087-1091(1992).
 CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
 CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR
 CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
 CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
 CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X59712; CAA42232.1;
 CC HSP; P19267; Ibfm.
 CC InterPro: IPR000947;
 CC Pfam: PF00808; CBFD_NFYB_HME; 1.
 CC PRINTS; PR00615; CCAATSUBUNTA.
 CC PROSITE; PS00685; CBFA_NFYB; 1.
 CC Transcription regulation; DNA-binding; Activator; Nuclear protein.
 CC DOMAIN 1 53 A DOMAIN.
 CC FT DOMAIN 54 143 B DOMAIN.
 CC FT DOMAIN 144 209 C DOMAIN.
 CC FT DNA_BIND 60 66 BY SIMILARITY.

SQ SEQUENCE 209 AA; 22676 MW; E89B0F0289882350 CRC64;
 Query Match 21.5%; Score 325.5; DB 1; Length 209;
 Best Local Similarity 53.2%; Pred. No. 7e-15;
 Matches 66; Conservative 18; Mismatches 37; Indels 3; Gaps 1;

QY 11 GAENGSAAGGANNAGAAQAAPAIREDRLMPIANVIRIMRVLPAHAHAKISDDAKETIO 70
 DB 32 GDAEGSLASGDHDESCGSK---DPYREQDIYLPANVARIMKTSIPSSGKIADAKECVQ 88
 QY 71 ECVSEYISFITEANERQOREKTTITADVLWAMSLRGDDYVEPLGAYLHRYREFEGD 130
 DB 89 ECVSEYISFITEANERQOREKTTITADVLWAMSLRGDDYVEPLGAYLHRYREFEGD 148
 QY 131 ARGV 134
 DB 149 EKI 152

RESULT 3
 CBFA_HUMAN
 ID CBFA_HUMAN STANDARD; PRT; 207 AA.
 AC P25208;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NUCLEAR TRANSCRIPTION FACTOR Y SUBUNIT BETA (NF-Y PROTEIN CHAIN B)
 DE (NF-YB) (CAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A) (CBF-A) (CAAT-
 DE BOX DNA BINDING PROTEIN SUBUNIT B).
 GN NFYB OR HAP3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92250488; PubMed=1577736;
 RA Li X.-Y., van Huijsduijn R., Mantovani R., Benoist C.O.,
 RA Mathis D.;
 RT "Intron-exon organization of the NF-Y genes. Tissue-specific splicing
 RT modifies an activation domain."
 RL J. Biol. Chem. 267:8984-8990(1992).
 CC [2]
 CC SEQUENCE FROM N.A.
 CC Badley Clarke J., Ting J.P.Y.;
 CC Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
 CC [3]
 CC SEQUENCE FROM N.A.
 CC MEDLINE=92195809; PubMed=1549471;
 CC Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I.,
 CC Benoist C., Mathis D.;
 CC "Evolutionary variation of the CCAAT-binding transcription factor
 CC NF-Y.";
 CC Nucleic Acids Res. 20:1087-1091(1992).
 CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
 CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR
 CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
 CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
 CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC EMBL; L06145; AA59930.1; -.
DR EMBL; X59710; CAA42230.1; ALT_INIT.
DR PIR; S22817; S22817.
DR HSP; P19267; 1BFM.
DR TRANSFAC; T00154; -.
DR MIM; 189904; -.
DR InterPro; IPR000947; -.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PS00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 1 52 A DOMAIN.
FT DOMAIN 53 142 B DOMAIN.
FT DOMAIN 143 207 C DOMAIN.
FT DNA_BIND 59 65 BY SIMILARITY.
SQ SEQUENCE 207 AA; 22831 MW; 1ADFA0E45F3CF972 CRC64;

Query Match 21.4%; Score 324.5; DB 1; Length 207;
Best Local Similarity 56.6%; Pred. No. 8.1e-15;
Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIREQDRMLPIANVIRIMRRVLPAHAKISDDAKETIQECVSEVISITGEANRCOREQR 93
Db : |||| : |||| : || : || : |||| : |||| : |||| : |||| : |||| : || :
51 SFREQDYLPIANVARIMKNAIPQTGKIADKCEVQECVSEFISFITSEASERCHQEKR 110
QY 94 KTTAEDVLWAMSRGLGDDYVEPLGAYLHRYREFEGDARGVGLVPGAAPSRGG 146
Db ||| : || : |||| : |||| : || : || : || : || : || : || : || : || :
111 KTINGEDILFAMSTGLGDSVVEPLKYLQKFRAMKRGKIG--GAVTATDG 160

RESULT 4
CBFA_MOUSE
ID CBFA_MOUSE STANDARD; PRT; 207 AA.
AC P22569;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN
DE CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B).
GN NFYB.
OS Mus musculus (Mouse), and Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090, 10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC SPECIES=Mouse;
RX MEDLINE=91006004; PubMed=1698608;
RA Hooft van Huijsduijnen R., Li X.-Y., Black D., Matthes H., Benoist C.,
RA Mathis D.;
RT "Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-
RT 1/CBF) subunits.";
RL EMBO J. 9:3119-3127(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse;
RX MEDLINE=92195809; PubMed=1549471;
RA Li X.-Y., Mantovani R., Hooft van Huijsduijnen R., Andre I.,
RA Benoist C., Mathis D.;
RT "Evolutionary variation of the CCAAT-binding transcription factor
RT NF-Y.";
RL Nucleic Acids Res. 20:1087-1091(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=Mouse;
RX MEDLINE=92250488; PubMed=1577736;
RA Li X.-Y., Hooft van Huijsduijnen R., Mantovani R., Benoist C.,
RA Mathis D.;
RT "Intron-exon organization of the NF-Y genes. Tissue-specific splicing
RT modifies an activation domain.";
RL J. Biol. Chem. 267:8984-8990(1992).

```

```

RN [4]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC SPECIES=Mouse;
RX MEDLINE=91093096; PubMed=2266139;
RA Vuorio T., Maiti S.N., de Crombrughe B.;
RT "Purification and molecular cloning of the 'A' chain of a rat
RT heteromeric CCAAT-binding protein. Sequence identity with the yeast
RT HAP3 transcription factor.";
RL J. Biol. Chem. 265:22480-22486(1990).
CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS. FOR
CC EXAMPLE IN TYPE 1 COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: TWO VARIANT SEQUENCES ARE THOUGHT TO BE
CC DERIVED FROM THE SAME GENE BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X55316; CAA39024.1; -.
DR EMBL; M55045; AAA40887.1; -.
DR EMBL; A20553; CAA01509.1; -.
DR PIR; S12045; S12045.
DR PIR; A23692; A23692.
DR PIR; F38245; F38245.
DR HSP; P19267; 1BFM.
DR TRANSFAC; T00083; -.
DR TRANSFAC; T00087; -.
DR TRANSFAC; T00613; -.
DR TRANSFAC; T00616; -.
DR MGD; MGI:97317; Nfyb.
DR InterPro; IPR000947; -.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PS00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 1 52 A DOMAIN.
FT DOMAIN 53 142 B DOMAIN.
FT DOMAIN 143 207 C DOMAIN.
FT DNA_BIND 59 65 BY SIMILARITY.
SQ SEQUENCE 207 AA; 22787 MW; 1ADFB4B04A3CFC22 CRC64;

Query Match 21.4%; Score 324.5; DB 1; Length 207;
Best Local Similarity 56.6%; Pred. No. 8.1e-15;
Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIREQDRMLPIANVIRIMRRVLPAHAKISDDAKETIQECVSEVISITGEANRCOREQR 93
Db : |||| : |||| : || : || : |||| : |||| : |||| : |||| : |||| : || :
51 SFREQDYLPIANVARIMKNAIPQTGKIADKCEVQECVSEFISFITSEASERCHQEKR 110
QY 94 KTTAEDVLWAMSRGLGDDYVEPLGAYLHRYREFEGDARGVGLVPGAAPSRGG 146
Db ||| : || : |||| : |||| : || : || : || : || : || : || : || : || :
111 KTINGEDILFAMSTGLGDSVVEPLKYLQKFRAMKRGKIG--GAVSATDG 160

RESULT 5
CBFA_CHICK
ID CBFA_CHICK STANDARD; PRT; 151 AA.
AC P25207;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)

```

DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN
DE CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B) (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92195809; PubMed=1549471;
RA Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I.,
RA Benoist C., Mathis D.;
RT "Evolutionary variation of the CCAAT-binding transcription factor
RT NF-Y";
RL Nucleic Acids Res. 20:1087-1091(1992).
CC -!- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
CC RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS, FOR
CC EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
CC -!- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
CC COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59713; CAA42233.1; ALT_SEQ.
DR HSSP: P19267; IBFM.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBF_D_NFYB_HMF; 1.
DR PROSITE: PS00685; CBF_A_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT DOMAIN 1 50 A DOMAIN.
FT DOMAIN 51 140 B DOMAIN.
FT DOMAIN 141 >151 C DOMAIN.
FT NON_TER 57 63 BY SIMILARITY.
FT NON_TER 151 151
SQ SEQUENCE 151 AA; 16733 MW; 3C2703043DE1A444 CRC64;

Query Match 21.3%; Score 322; DB 1; Length 151;
Best Local Similarity 59.8%; Pred. No. 8.6e-15;
Matches 61; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 34 AIRQDRLMPINIRIMRVLPANAKISDDAKETIQECVSEYISFITGEANERCQEQ 93
Db 49 SFREQDIYLIPIANVARIMKNAIQTKIAKAEQVCEVSEFISITSEASERCHQER 108

QY 94 KTTAEDVLWAMSRIGFDYVEPLGAYLHRYREFEGDARGVG 135
Db 109 KTINGEDILFAMSTGLGSDSYVEPLKLYLQKFRAMKRGKIG 150

RESULT 6
ID HAP3_YEAST STANDARD; PRT; 144 AA.
AC P13434;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRANSCRIPTIONAL ACTIVATOR HAP3 (UAS2 REGULATORY PROTEIN A).
GN HAP3 OR YBL021C OR YBL0441.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BMG1-7A;
RX MEDLINE=88174707; PubMed=2832732;
RA Hann S., Pinkham J., Wei R., Miller R., Guarente L.;
RT "The HAP3 regulatory locus of Saccharomyces cerevisiae encodes
RT divergent overlapping transcripts";
RL Mol. Cell. Biol. 8:655-663(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=94103216; PubMed=8276800;
RA van Dyck L., Pearce D.A., Sherman F.;
RT "PIM1 encodes a mitochondrial ATP-dependent protease that is required
RT for mitochondrial function in the yeast Saccharomyces cerevisiae";
RL J. Biol. Chem. 269:238-242(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RA Goffeau A., Jonniaux J.-L., Purnelle B., Skala J., de Wergifosse P.,
RL Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP MUTAGENESIS.
RX MEDLINE=94038948; PubMed=8223474;
RA Xing Y., Fikes J.D., Guarente L.;
RT "Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding
RT domain";
RL EMBO J. 12:4647-4655(1993).
CC -!- FUNCTION: BINDS TO THE UPSTREAM ACTIVATION SITE (UAS) OF THE CYCL
CC GENE AND OTHER GENES INVOLVED IN MITOCHONDRIAL ELECTRON TRANSPORT
CC AND ACTIVATES THEIR EXPRESSION. RECOGNIZES THE SEQUENCE CCAAT.
CC -!- SUBUNIT: TWO COMPLEXES BIND CCAAT; COMPLEX I, THAT CONSISTS OF
CC HAP2/3/5 AND COMPLEX II, THAT CONSISTS OF HAP2/3/5/4.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M20318; AAA53538.1; -.
DR EMBL: X74544; CAA52633.1; -.
DR EMBL: Z35782; CAA84840.1; -.
DR PIR: A28123; A28123.
DR TRANSFAC: T00350; -.
DR SGD: S0000117; HAP3.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBF_D_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNITA.
DR PROSITE: PS00685; CBF_A_NFYB; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 1 35 A DOMAIN.
FT DOMAIN 36 125 B DOMAIN.
FT DOMAIN 126 144 C DOMAIN.
FT DOMAIN 144 48
FT DNA_BIND 42 48
FT DOMAIN 69 80
SQ SEQUENCE 144 AA; 16154 MW; 7AB5027BAE420337 CRC64;

Query Match 19.5%; Score 295.5; DB 1; Length 144;
Best Local Similarity 50.4%; Pred. No. 4.3e-13;
Matches 60; Conservative 21; Mismatches 33; Indels 5; Gaps 2;

QY 8 PAAGAENSAAGGANGGAQOHAAPAIRODRMLPTIANVIRMRRLPAHAHAKISDDAKE 67
 Db 13 PEDTQENG---GNASSGSLQ---ISTIREQDRWLPINNVARLMKNTLPPSAKVSXDAKE 67
 QY 68 TIQCEVSEYISFITGEANERCOREORKTITAEDVWAMSRGLGDDYVEPLGAYLHYRE 126
 Db 68 CMQCEVSELISFVTSASDRCADKRTINGEDILISLHALGFENYAEVLKIYLAQRQ 126

RESULT 7
 ID PHP3_SCHPO STANDARD; PRT; 116 AA.
 AC P36611;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE TRANSCRIPTIONAL ACTIVATOR PHP3.
 GN PHP3 OR SPAC23C11.08.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=401;
 RX MEDLINE=94038948; PubMed=8223474;
 RA Xing Y., Fikes J.D., Guarente L.;
 RT "Mutations in yeast HAP2/HAP3 define a hybrid CCAAT box binding
 RT domain.";
 RL EMBO J. 12:4647-4655(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Brown D., Churche C.M., Barrell B.G., Rajandream M.A., Wood V.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: BELONGS TO A COMPLEX THAT BINDS TO THE SEQUENCE CCAAT
 CC LOCATED UPSTREAM OF GENES INVOLVED IN MITOCHONDRIAL ELECTRON
 CC TRANSPORT.
 CC -!- SUBUNIT: BELONGS TO AN HETEROTRIMERIC CCAAT-BINDING COMPLEX (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X75072; CAA52966.1; -;
 DR EMBL: 298559; CAB11161.1; -;
 DR PIR: S42744; S42744.
 DR HSSP: P19267; 1BPM.
 DR TRANSFAC: T01263; -;
 DR InterPro: IPR0000947; -;
 DR Pfam: PF00808; CBF_NFYB_HMF; 1.
 DR PRINTS: PR00615; CCAATSUBUNTA.
 DR PROSITE: PS00685; CBF_NFYB; 1.
 DR KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
 FT DOMAIN 1 6 A DOMAIN.
 FT DOMAIN 7 95 B DOMAIN.
 FT DOMAIN 96 116 C DOMAIN.
 FT DNA_BIND 12 18 BY SIMILARITY.
 FT SEQUENCE 116 AA; 12909 MW; 930DFCA773099CB5 CRC64;

Query Match

18.1%; Score 274; DB 1; Length 116;

Best Local Similarity 60.5%; Pred. No. 8.5e-12;
 Matches 52; Conservative 18; Mismatches 16; Indels 0; Gaps 0;

QY 41 LMPIANVIRMRRLPAHAHAKISDDAKETIQCEVSEYISFITGEANERCOREORKTITAE 100
 Db 11 LPIANVARIMRSALPENAKISKAKDCVQDCVSEFISFTVGEASEQCTQERKKTITGD 70
 QY 101 VLWAMSRGLGDDYVEPLGAYLHYRE 126
 Db 71 VLLALNTLGFENYAEVLKISLTKYRE 96

RESULT 8
 ID HAP3_KLULA STANDARD; PRT; 205 AA.
 AC P40914;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 DE HAP3 TRANSCRIPTIONAL ACTIVATOR.
 GN HAP3.
 OS Kluyveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140;
 RX MEDLINE=95147853; PubMed=7845362;
 RA Mulder W., Scholten I.H.J.M., de Boer R.W., Grivell L.A.;
 RT "Sequence of the HAP3 transcription factor of Kluyveromyces lactis
 RT predicts the presence of a novel 4-cysteine zinc-finger motif.";
 RL Mol. Gen. Genet. 245:96-106(1994).
 CC -!- FUNCTION: BINDS TO THE UPSTREAM ACTIVATION SITE (UAS) OF THE CYC1
 CC GENE AND OTHER GENES INVOLVED IN MITOCHONDRIAL ELECTRON TRANSPORT
 CC AND ACTIVATES THEIR EXPRESSION. RECOGNIZES THE SEQUENCE CCAAT.
 CC -!- SUBUNIT: BELONGS TO AN HETEROTRIMERIC CCAAT-BINDING COMPLEX (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.
 CC -!- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
 CC DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
 CC SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L25779; AAC41662.1; -;
 DR PIR: S41926; S41926.
 DR InterPro: IPR0000947; -;
 DR Pfam: PF00808; CBF_NFYB_HMF; 1.
 DR PRINTS: PR00615; CCAATSUBUNTA.
 DR PROSITE: PS00685; CBF_NFYB; 1.
 DR KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
 FT DOMAIN 1 20 A DOMAIN.
 FT DOMAIN 21 110 B DOMAIN.
 FT DOMAIN 111 205 C DOMAIN.
 FT DNA_BIND 27 33 BY SIMILARITY.
 FT DOMAIN 54 65 IMPORTANT FOR SUBUNIT INTERACTIONS
 FT FT (BY SIMILARITY).
 FT DOMAIN 124 149 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 156 179 SER/THR-RICH.
 FT SEQUENCE 205 AA; 23122 MW; B33C7B6FBELD7E1A CRC64;

Query Match

17.5%; Score 265; DB 1; Length 205;

Best Local Similarity 54.3%; Pred. No. 5.7e-11;
 Matches 50; Conservative 17; Mismatches 25; Indels 0; Gaps 0;

QY 35 IREQDRLMPIANVIRMRRLPAHAKISDDAKETIOECVSEVISFITGEANERCOREQK 94
Db 20 LABQDRWLPINNVRUMKMTLPATTKVSKDAKCEQCEVSEFISVETSEACDRCTSGRRK 79
QY 95 TITAEVDLWMSRLGDFDDVPEPLGAYLHYRE 126
Db 80 TINGEDILLSLHUALGEPENAEVVKIYLAKYQ 111
RESULT 9
ID CBFA_XENLA STANDARD; PRT; 122 AA.
AC P25211;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A) (NF-Y PROTEIN
CHAIN B) (NF-YB) (CAAT-BOX DNA BINDING PROTEIN SUBUNIT B) (FRAGMENT).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92195809; PubMed=1549471;
RA Li X.-Y., Mantovani R., Hooft van Huijsduijn R., Andre I.,
RA Benoist C., Mathis D.;
RT "Evolutionary variation of the CCAAT-binding transcription factor
NF-Y";
RL Nucleic Acids Res. 20:1087-1091(1992).
CC -1- FUNCTION: STIMULATES THE TRANSCRIPTION OF VARIOUS GENES BY
RECOGNIZING AND BINDING TO A CCAAT MOTIF IN PROMOTERS. FOR
EXAMPLE IN TYPE I COLLAGEN, ALBUMIN AND BETA-ACTIN GENES.
CC -1- SUBUNIT: HETEROMERIC TRANSCRIPTION FACTOR COMPOSED OF TWO
COMPONENTS, A AND B, THAT ARE BOTH NEEDED FOR DNA BINDING.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DOMAIN: CAN BE DIVIDED INTO THREE DOMAINS: THE WEAKLY CONSERVED A
DOMAIN, THE HIGHLY CONSERVED B DOMAIN THOUGHT TO BE INVOLVED IN
SUBUNIT INTERACTION AND DNA BINDING, AND THE GLU-RICH C DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE CBF-A SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; X59709; CAA42229.1; -
DR PIR; S22819; S22819;
DR InterPro; IPR000947; -
DR PROSITE; PS00685; CBFA_NFYB; 1.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT NON_TER 1 1
FT DOMAIN <1 57 B DOMAIN.
FT DOMAIN 58 122 C DOMAIN.
FT SEQUENCE 122 AA; 13498 MW; E76AD9FBC638F53A CRC64;

Query Match 15.4%; Score 232.5; DB 1; Length 122;
Best Local Similarity 50.0%; Pred. No. 4.4e-09;
Matches 49; Conservative 14; Mismatches 32; Indels 3; Gaps 2;

QY 69 IOECVSEVISITGEANERCOREQKTTITAEVDLWMSRLGDFDDVPEPLGAYLHYREFE 128
Db 1 VOECVSEFISITSEASRCHQERKKTINGEDILLFAMSTIGFDSYVEPLKLYLOKFEAM 60

QY 129 GDARGVGLVPGAAPSRGGD--HHPH-SMSPAAMLSKRG 163
Db 61 KGEGIGCTVTGDLGELDAEPEFTSQIPAGLITTDG 98

RESULT 10
DRI_ARATH STANDARD; PRT; 159 AA.
ID DRI_ARATH
AC P49592;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE DRI PROTEIN HOMOLOG.
GN DRI.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV. COLUMBIA;
RX MEDLINE=95116318; PubMed=7816619;
RA Kuromori T., Yamamoto M.;
RT "Cloning of cDNAs from Arabidopsis thaliana that encode putative
protein phosphatase 2C and a human Dri-like protein by transformation
of a fission yeast mutant";
RL Nucleic Acids Res. 22:5296-5301(1994).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: TO HUMAN PROTEIN DRI.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; D38110; BAA07288.1; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
KW Transcription; Nuclear protein.
SQ SEQUENCE 159 AA; 17942 MW; B5BCEC9E9BBF54FE CRC64;

Query Match 12.9%; Score 195; DB 1; Length 159;
Best Local Similarity 28.0%; Pred. No. 1.5e-06;
Matches 46; Conservative 32; Mismatches 60; Indels 26; Gaps 4;

QY 37 EODRLMPIANVIRMRRLPAHAKISDDAKETIOECVSEVISFITGEANERCOREQKTI 96
Db 11 KEDASLPKATMTKIIEMLPPDVRVARDAQDILLIECCVEFINLVSSNDVCKEDKRTI 70

QY 97 TAEDVLWMSRLGDFDDVPEPL-GAY-LHYREFEGDARGVGLVPGAAPSRGDDHHPHMS 154
Db 71 APEHVLKALQVLGFGYEIEVYAAVEQHKYETMDTQSRVKNPG----- 115

QY 155 PAAMLKSRGVSAAAMLPHHHHHHDMQMAAMYGTAVPVPPAGP 198
Db 116 -AQMTTEEAABEQQRMF-----AARARMNGGVSVPOPEHP 150

RESULT 11
ID MAF2_RAT STANDARD; PRT; 369 AA.
AC P54844;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR MAF2 (PROTO-ONCOGENE C-MAF).
GN MAF2 OR MAF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;


```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=WISTAR;
RX  MEDLINE=97190228; PubMed=9038383;
RA  Sakai M., Imaki J., Yoshida K., Ogata A., Matsushima-Hibaya Y.,
RA  Kuboki Y., Nishizawa M., Nishi S.;
RT  "Rat maf related genes: specific expression in chondrocytes, lens and
RT  spinal cord.";
RL  Oncogene 14:745-750(1997).
CC  -!- SUBCELLULAR LOCATION: NUCLEAR.
CC  -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC  -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
DR  EMBL; U56242; AAB50063.1; -.
DR  HSP; P05412; IJUN.
KW  Proto-oncogene; Transcription regulation; DNA-binding; Activator;
KW  Nuclear protein.
FT  DOMAIN 139 146 POLY-ALA.
FT  DOMAIN 169 173 POLY-ALA.
FT  DOMAIN 180 187 POLY-HIS.
FT  DOMAIN 191 194 POLY-HIS.
FT  DOMAIN 212 220 POLY-GLY.
FT  DOMAIN 225 234 POLY-GLY.
FT  DOMAIN 284 310 BASIC MOTIF.
FT  DNA_BIND 312 333 LEUCINE-ZIPPER.
FT  DOMAIN 369 AA; 288E464708DA6C7D CRC64;
SQ  SEQUENCE 369 AA; 38457 MW; 288E464708DA6C7D CRC64;

Query Match 11.8%; Score 179; DB 1; Length 369;
Best Local Similarity 30.9%; Pred. No. 3.8e-05;
Matches 60; Conservative 15; Mismatches 63; Indels 56; Gaps 10;

QY 99 EDVLWAM-----SRLGF--DDVVEPLGAYLHRYR-EFEGDARGVGLV-----PGAAP 142
DB 89 EDYWMYTGYPQQLNPEALGFSPEDAVEALISNSHQLGGFDGYARGAQQLAAAGAGAGA 148
QY 143 SRGDDHHPHSMSPA-----AMLKSRGPVSGAAMLPHHHHHHDMHOMHAAMYGTTAVPPAG 199
DB 149 SLGGS--GEEMGPAAVVSAVIAAAASGAGAPHYHHH-----185
QY 200 HHGGLFPHPGQSSHYLPYAYEPTYGGEHAMAAYTYGGAAYAPGNGSGSGSGGGGSA 259
DB 186 HH-----ATGHH-----HHPTAGAPCAAGSASASAGAGAGGPPASAGGGGGGG 232
QY 260 SHTPQSGGLEHHPH 273
DB 233 GGT-AGAGGALHPH 245

RESULT 12
MAF2_MOUSE
ID MAF2_MOUSE STANDARD; PRT; 370 AA.
AC P54843;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION FACTOR MAF2 (PROTO-ONCOGENE C-MAF).
GN MAF2 OR MAF.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BAUB/C; TISSUE=Cerebellum;
RX MEDLINE=95097997; PubMed=7799931;

```

```

RA Kurschner C., Morgan J.I.;
RT "The maf proto-oncogene stimulates transcription from multiple sites
RT in a promoter that directs purkinje neuron-specific gene
RT expression.";
RL Mol. Cell. Biol. 15:246-254(1995).
CC -!- FUNCTION: THE C-MAF INTERACTION SITE WAS MAPPED TO THE SEQUENCE
CC 5'-[GTG[GC]N[GT]NCTCAGN-3' IN THE L7 PROMOTER. IT MAY INTERACT
CC WITH ADDITIONAL BASIC-ZIPPER PROTEINS THAT DETERMINE A SUBTYPE OF
CC MAF-RESPONSIVE ELEMENT BINDING.
CC -!- SUBUNIT: HOMO- OR HETERODIMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; S74567; AAB32820.1; -.
DR HSP; P05412; IJUN.
DR TRANSFAC; T01432; -.
DR MGD; MGI:96909; Maf.
KW Proto-oncogene; Transcription regulation; DNA-binding; Activator;
KW Nuclear protein.
FT DOMAIN 139 146 POLY-ALA.
FT DOMAIN 169 173 POLY-ALA.
FT DOMAIN 180 187 POLY-HIS.
FT DOMAIN 191 194 POLY-HIS.
FT DOMAIN 207 210 POLY-SER.
FT DOMAIN 212 220 POLY-GLY.
FT DOMAIN 225 235 POLY-GLY.
FT DNA_BIND 285 311 BASIC MOTIF.
FT DOMAIN 313 334 LEUCINE-ZIPPER.
SQ SEQUENCE 370 AA; 38655 MW; 76A9517EFF9C777C CRC64;

Query Match 11.7%; Score 176.5; DB 1; Length 370;
Best Local Similarity 33.5%; Pred. No. 5.6e-05;
Matches 67; Conservative 15; Mismatches 51; Indels 67; Gaps 14;

QY 99 EDVLWAM-----SRLGF--DDVVEPLGAYLHRYR-EFEGDARGVGLV-----PGAAP 142
DB 89 EDYWMYTGYPQQLNPEALGFSPEDAVEALISNSHQLGGFDGYARGAQQLAAAGAGAGA 148
QY 143 SRGDDHHPHSMSPA-----AMLKSRGPVSGAAMLPHHHHHHDMHOMHAAMYGTTAVPPAG 197
DB 149 SLGGS--GEEMGPAAVVSAVIAAAASGAA--PHYHHHHH---HA-----AG 190
QY 198 PPHGGFLPHPGQSSHYLPYAYEP-TYGGEHAMAAYTYGGAAYAPGNGG---SDGSGSG 253
DB 191 HHHH-----PTAGAPGTAGGASSSSNGAGGA---GGGGPANTGGGGGD 231
QY 254 GGGGSASHTPQSGGLEHHPH 273
DB 232 GGGGTA-----GAGGALHPH 246

RESULT 13
HMDP_DROME
ID HMDP_DROME STANDARD; PRT; 590 AA.
AC P07548;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 04-NOV-1995 (Rel. 32, Last annotation update)
DE HOMEOTIC DEFORMED PROTEIN.
GN DFD.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```

QY	156	AAMLSKRGVPVGAAMLPHHH-----HHHDMQM-----HAAMYGG-----TA	191
Db	83	ADMVSD-----YMAHHNHPHSHSHSTHSLPHHHNSAISGHQASAGVSSNYANA	134
QY	192	VPP--PAGPPHHGGFLMHP-OGSSHYLPYAYE-----PTYGGEHAMAAYGGAA	238
Db	135	TPSPHPSHPH-----AHPQSLGYVVHHAPEFISAGAVHSDPTNG-----YGPAA	180
QY	239	YAP--GNGGSGDGS--GGG--GGSAS-----HTPQSGGGLXH-----PH	273
Db	181	NVPNTSNGGGGGSGAVLGGAVGSSANGYGYGGYGGTANGSVGTHSQGHS	236
RESULT 14			
ID	TMAF_AVIS4	STANDARD; PRT; 369 AA.	
ID	TMAF_AVIS4	STANDARD; PRT; 369 AA.	
AC	F23091;		
DT	01-NOV-1991 (Rel. 20, Created)		
DT	01-NOV-1991 (Rel. 20, Last sequence update)		
DT	30-MAY-2000 (Rel. 39, Last annotation update)		
DE	TRANSFORMING PROTEIN MAF.		
GN	V-MAF.		
OS	Avian musculoaponeurotic fibrosarcoma virus AS42.		
OC	Viruses; Retrod viruses; Retroviridae; Avian type C retroviruses.		
OX	NCBI_TaxID=11873;		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=90046665; Pubmed=2554284;		
RT	Nishizawa M., Kataoka K., Goto N., Fujiwara K.T., Kawai S.;		
RT	"v-maf, a viral oncogene that encodes a 'leucine zipper' motif.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7711-7715(1989).		
CC	-1- FUNCTION: MIGHT BE A TRANSCRIPTIONAL TRANS-ACTIVATOR.		
CC	-1- SUBCELLULAR LOCATION: NUCLEAR.		
CC	-1- DISEASE: INDUCES MUSCULOAPONEUROTIC FIBROSARCOMA IN CHICKENS.		
CC	-1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A ENV-MAF		
CC	POLYPROTEIN.		
CC	-1- SIMILARITY: BELONGS TO THE BZIP FAMILY. MAF SUBFAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announcement/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; M26769; AAA42377.1; --		
DR	PIR; B33975; TVFVAF.		
DR	HSP; P05412; 10UN.		
DR	TRANSFAC; T01430; --		
KW	Transferring protein; Transcription regulation; DNA-binding;		
KW	Nuclear protein.		
FT	DOMAIN 163		
FT	DOMAIN 173		
FT	DOMAIN 182		
FT	DOMAIN 184		
FT	DOMAIN 214		
FT	DOMAIN 223		
FT	DOMAIN 233		
FT	DOMAIN 274		
FT	DNA_BIND		
FT	DOMAIN 302		
FT	LEUCINE-ZIPPER.		
FT	SEQUENCE 369 AA; 38892 MW; F368B220ACE50FF6 CRC64;		
QY	99	EDVLWAM-----SRIGF--DDYVEPL-----GAYLHRYREFEGDARGVGLVPG	139
Db	89	EDYYWMTGYPQOLNPEALGFSPEDAVEALINSSHHPLPCA-----PDGYARGQQL----	138
QY	140	AAPSRGGDHHPSHSGPAAMLSKRGVPSGAAM--LPHHHHHHHDMQMAHYGGTAVPPAG	197
Db	140	AAPSRGGDHHPSHSGPAAMLSKRGVPSGAAM--LPHHHHHHHDMQMAHYGGTAVPPAG	197
Query Match 11.28; Score 170; DB 1; Length 369;			
Best Local Similarity 31.18; Pred. No. 0.00015;			
Matches 61; Conservative 9; Mismatches 52; Indels 74; Gaps 11			

Db 139 -AAAAGGSVPAEEMGSAAAVVAVIAAAAAAGGAPHYHHHH----- 179
Qy 198 PPHGGFLMHPHOGSSHYLPYAYEPTYGGSHMAAAYGGAAYAPGNGSGDGGSGGGG 257
Db 180 HPHGG-----GGG-----GGGHPGCAAPGS---APPSSASSAAGSGGGG 218
Qy 258 SASHTPOGSGGLEPH 273
Db 219 GGG---GGAGGLHHPH 231

RESULT 15
TBAP_HUMAN STANDARD; PRT; 176 AA.
AC Q01658;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TATA-BINDING PROTEIN-ASSOCIATED PHOSPHOPROTEIN (DOWN-REGULATOR OF
DE TRANSCRIPTION 1) (DRI PROTEIN).
GN DRI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=92354065; PubMed=1339312;
RA Inostroza J.A., Mermelstein F.H., Ha I., Lane W.S., Reinberg D.;
RT "Dri, a TATA-binding protein-associated phosphoprotein and inhibitor
RL of class II gene transcription.";
Cell 70:477-489(1992).
CC -|- FUNCTION: THE ASSOCIATION OF DRI WITH TBP RESULTS IN A FUNCTIONAL
CC REPRESSION OF BOTH ACTIVATED AND BASAL TRANSCRIPTION OF CLASS II
CC GENES. THE INTERACTION OF DRI WITH TBP PRECLUDES THE FORMATION OF
CC A TRANSCRIPTION-COMPETENT COMPLEX BY INHIBITING THE ASSOCIATION OF
CC TFIIA AND/OR TFIIIB WITH TBP.
CC -|- SUBUNIT: DRI EXISTS IN SOLUTION AS A HOMOTETRAMER THAT DISSOCIATES
CC DURING INTERACTION WITH TBP AND THEN, AFTER COMPLEXING WITH TBP,
CC REASSOCIATES AT A SLOW RATE, TO RECONSTITUTE THE TETRAMER.
CC -|- SUBCELLULAR LOCATION: NUCLEAR.
CC -|- PTM: PHOSPHORYLATION REGULATES ITS INTERACTION WITH TBP.
CC -|- SIMILARITY: TO A THALIANA DRI HOMOLOG AND TO YEAST HAP3 PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M97388; AAA5842.1; -;
DR PIR: A43320; A43320.
DR TRANSFAC: T00197; -;
DR MIM: 601482; -;
DR InterPro: IPR000947; -;
DR Pfam: PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
KW Transcription; Phosphorylation; Nuclear protein.
FT SIMILAR 1 100 SOME, TO YEAST HAP3 PROTEIN (AA 30-131).
FT DOMAIN 100 103 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 121 168 ALA/GLN-RICH.
SQ SEQUENCE 176 AA; 19443 MW; 36E7E59F2FD6CAB5 CRC64;

Query Match 10.6%; Score 161; DB 1; Length 176;
Best Local Similarity 30.7%; Pred. No. 0.00027;
Matches 35; Conservative 30; Mismatches 41; Indels 8; Gaps 2;
Qy 31 AAPAIREQDRMLPTANVIRMRVLPAAHAKISDDAKETIQECVSEYISFTGEANERCQR 90
Db 2 ASSGNDLDTIPRAAINKMIKETLP-NVRVANDARELVVNCCTEFTHLISSEANEICNK 60

Qy 91 EQKRTITAEEDVLWAMSRQLGDDYVEPLGAYLHRYREFECGARGVGLVPCAAPS 144
Db 61 SEKKTISPHEHVIALESGLF-----GSYISEYKEVLQECKTVALKRRKASSR 107

Search completed: October 25, 2001, 09:24:50
Job time: 185 sec

2834..2932,3028..3114,3222..3365)
/note="gene_id:MNJ7.1"
/codon_start=1
/pseudo

/evidence=not_experimental
/product="formyltetrahydrofolate deformylase-like"
complement(join(3514..3755,3836..3893,3983..4032,
4110..4289,4380..4773,4852..5148))
/note="gene_id:MNJ7.2
pirilG17442

strong similarity to unknown protein"

CDS

/codon_start=1
/evidence=not_experimental
/protein_id="BAB09069.1"
/db_xref="GI:9758771"
/translation="MEGGFYSDWNSSSSLFGSENPHELEGNVRSEETVSOIPOPO
TPREPMEFLRSWSLSASEISKALAQKQRODLFSVSONSPRGFEQDVAADPLMAEN
IMNSAGTRRSGLSKWFKHKOHTNPSTIRPKRKARVOKAHVHSAVSTAALAGLA
SVTSEESKESCSMALASATELLASHCIDMAEOAGADHTCVASTVRSXVDIHSIP
GDLMTLAAATALRGAALKVRQPKSRKNATITPCERSFSDSHMPGENCQRLDEP
NLPLEGLVOCARNGLRKRVCIYINKSQVMIKLKSKHVGGAFAKIKCVVYGVD
EISAPCRKRENSEEVYFGLTQGLLEFKCKSKIQKQWVAGIQSNLRVLSCLEAA
KCSLESLSLRNM"

complement(6607..7680)
/note="contains similarity to retroelement pol
polyprotein

CDS

/codon_start=1
/evidence=not_experimental
/protein_id="BAB09070.1"
/db_xref="GI:9758772"
/translation="MVSVEKLRRTRKSKISAEKTPDRTQSSPIPPVPISSPIQON
KATFVVDGTLPRPEPDPSPVMSRCNSVMKSLVSVTKQIYKSLRNDAAETAK
DLDRPHNLPRSYQLTQIWSLQOQNSLSIDYTYTLTLDDLDGASCVNCRCK
CCSATVSNHKSIVKLEAGLNFYTSQIKMKKTLAEIYNLLDQDHQSRNIV
TPTNNAFNVSLISNDQIAVNATRSFQNPSPKQVCSHCYGTGHTADTCYKHGPI
GFKHKQVHHLRQNRLLTL"
complement(join(8453..8827,9033..9410))
/note="gene_id:MNJ7.4"
/codon_start=1
/evidence=not_experimental
/product="membrane channel
(tonoplast intrinsic protein)-like"
/protein_id="BAB09071.1"
/db_xref="GI:9758773"
/translation="MVKTEVSGVDSFSVSSLKAYLSSEFIATLLFVAGVGSVAPAK
LTSGLDPAGLVAIAAHAFALFVGVSIANISGHLNPAVTLGLAIGNLTITGF
FYTAQCGSIVACLLVFTNGKSVPTHGVSAGLGAEGVYMEIVTFALVIVYAT
AADPKGSLGTPIATIGFIVGANILAAFPFGSGMNPARSFGPAVVSGLSQIWIY
VGPIVGGALAGLIYGVFISYEAVETREIV"
complement(12407..14137)
/note="gene_id:MNJ7.5"
/codon_start=1
/evidence=not_experimental
/product="selenium-binding protein-like"
/protein_id="BAB09072.1"
/db_xref="GI:9758774"
/translation="MLRTVSNFTTRSHVSGTASSNSWTIVPALRFGTGLVRAAV
ELINDGEPRDASPLVHLRVSNGVYVSLCQLHGTYTKGFSVNTSLMSLFYKT
SDSLDAHKVDEMPDPDPIVSNLSVGVQSGRPFQEGCLFLLEHRSDFPNFEST
ALAAACARLHSPGLGACVTHSKVLGKGVVYVGNCLIDMYKCGFMDADVLFQHM
BEKTVSNVATVASCNRKLELGLFWFFHOMPDPDTVTYNELIDAFYKGFDPNNAFQV
LWMPNPNSSWNTLTGTVVNSEKSGEATEFTKMHSSGVRFDEYSISIVLAAVAALA
VPMGSLIHACAKLGLDSRVVASALIDMYSKCMLKHAELMFTWPKRNLIVNWM
ISGYARNDSIEAKLNFQLQERFLKPDRETFNLNLAVCSEVPMVLMGYEMIT
NEYTRKPSVHCCLSRIMGORGVEWQAKQVIFRGVGDVGNRALLGACSRKDLK
RAKTVAAKMIELGADKDEYLIYVNSNLYAYHERVGVQIRKIRRESGVGSSW
IDSRPKCSSY"
join(14542..14786,14885..14950,15046..15162,15242..15464,
15527..15691,15784..15935,16196..16322)
/note="gene_id:MNJ7.6"
/codon_start=1

CDS

/evidence=not_experimental
/product="nodulin-like protein"
/protein_id="BAB09073.1"
/db_xref="GI:9758775"
/translation="MOFFCVNLYRSVLNLLERMKTEIEMVIVGGLVMVQFVYVAGN
SLMSYLSLGLGPTIVIFESFATFIILSPAILFERKOMPNEISLRLIGKLVLSF
AGVTLFQSLLEGIRLTSPAMATNPANLAPGLFFIANIVGLKKNLKVYSKLILG
TLICVFGALAMSVMIHSTISHKEDDTPIFVDRDKVGGCIYLLGAVFLVSTNVVLA
STLABFPAPISLSALTALLGLITTVLLLNQRKTKVLASSISFGNLGVYSVLGAV
SCAVSFNGWAMKKGKPVFVSFPAIVISVAFVLTGSESVSLGSGVGMVLMFVGL
YLVLWAKGKEGFESESEBFSKPKPLLS"
join(17364..19016,19220..19540,19635..19760,19841..19963,
20046..20282,20360..20494,20688..20815,20899..21007,
21259..21546,21762..21914,21995..22293,22596..23076)
/note="dbj|BAA90625.1
gene_id:MNJ7.7
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09074.1"
/db_xref="GI:9758776"
/translation="MASTADFLDDQTDDEDFDKLVDDSYTPTASSAKELKFDGSD
SDAKAFANLSVVDVLDGDDVALNEAGLNDVANEGTSGSVKKEPSSSIAPAVQF
VNSDANRLDQDVVRSEVDDMALTEGKESNIVDGGSGVKEVDMGSEYADSVNDG
GGFGSYDFTELDATAGNVQCAEAVATAGNLVANDTINTSVGLDSAGEHQGQ
VQHDGSGQYVDSQSSEWNLYPGWKYDASTGQYVQDQDATVMSQESTINSTGNWES
VAADNSVAYLKQSTTSAMAGTAEVSTWNOVSQVNGYPERHVEDAQYPGWYDTIA
QEWRLDSYNOASQTTVTQGAHQDQVGHARTTYYHNNSQSVYDVNNKQTKRAQD
FATOGHSGWDESYVANNQAGNTWQVNVGAKEPAVTSLSLRFGGNQVNNLYSTE
SVAEOPKNTIGNQSFIPQHMVASATONGPLSEFNLDLYNRQGSVDIAKQSPNNQLF
SPSVGRSSDRRPHALVSGGKLVLMKDNQNSLQNTSFGSQGGSGSITVNLAEV
ISGASYSYSGPGBDSLSYFRLHQHQLPGLVGNVSGSEKHLHWIDERLHLCSSNMDF
SRKLLKMLLSLRLTSCQYVYKLRSPGSDASOKETDPEAAVAKLFAFKKDGTONG
YAPISQCHLPESQOMQVTADEVQNLASGRKMEALCAQEGHILWGALVLAQGLD
QFVDTVKOMALROLTPGSLPLTCLLVAGQPAEVCPTGSSSLDNWENGLGITAN
RTDDDLVLIHLGDSMMKRGERTIAHICLYLADKNPDYSESARCLVGDHMKCPR
TASPDIAQTELYEYKSLNSQYILLFPQPKYIYAHMLAEVKLSAQYCCAVI
RCLKTSRSSEVEMKQFASSLEERSHQEGNLAPAKLVGKLLNSLWMPPPAPHSV
TGNPQYNEYHQHQEAALKLSQSANTMSLMPASIEPVHEMGNGRTMAAHSRVS
EPDFSKTPIQDDTDSKAPQGVQVTKRKVPSPSRFRFGIGILKNTYKGVFPFSRS
SNEAKLGNENQFYVDNLRKVRGVEKPAEAAALPPPTVPFRNSLGHENKSEIK
NEMSPSSGSSSGSPTPSPSPGIPVPSOGSOFSGARGMVRYVDTYNGSSSMY
QSPVSSRPPPIPAKAKFPVPAAPASFANDQVMESVSAETROENSGDEAVGSAAGP
PSQASHPTPSPPIAMQRPFPVDNIRRSQSGSLGDLFQSVSRRTASGSGSVNSSF
MSPTASATFRPSPNSSSSLGEEQVEL"
join(24619..26247,26400..26723,26819..26944,27044..27166,
27252..27422,27695..27820,27896..28030,28231..28358,
28444..28552,28759..29079,29155..29304,29383..29669,
29975..30431)
/note="contains similarity to unknown protein
dbj|BAA90625.1
gene_id:MNJ7.8"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB09075.1"
/db_xref="GI:9758777"
/translation="MASASQFLEDDQTDDEDFDKLVDDAYSPTAQASSVTELKFPDD
ESDSDITRAFNSLISIGDPIGGDGTLEAIIGNDVANEGASGSGVEDEPSSIAPEAV
QFPHSDARELDDMERSEVADMPLESEKACTIVNEPGIPGVKELDWGSDADLSVND
GRGFSYSDFTELDATAGNLQKADVAVATAGNLVANDTNTSVGEHQGLHDS
ASQOYVDSQSSEWNLYPGWKYDASTGQVQVGDHDSMNSQSEYENSTNWNVAANN
SDVYQRQSTASVAGTVENSTWNOVSQVSGYPRHMFVDSQYPGWYDTYINQEWRS
LDSYNAQFTQGAANDQVQNGNSFTAVDHSRESNVHDVYDKNQILRTKFDIQSOHG

CDS

/evidence=not_experimental
/product="membrane channel
(tonoplast intrinsic protein)-like"
/protein_id="BAB09071.1"
/db_xref="GI:9758773"
/translation="MVKTEVSGVDSFSVSSLKAYLSSEFIATLLFVAGVGSVAPAK
LTSGLDPAGLVAIAAHAFALFVGVSIANISGHLNPAVTLGLAIGNLTITGF
FYTAQCGSIVACLLVFTNGKSVPTHGVSAGLGAEGVYMEIVTFALVIVYAT
AADPKGSLGTPIATIGFIVGANILAAFPFGSGMNPARSFGPAVVSGLSQIWIY
VGPIVGGALAGLIYGVFISYEAVETREIV"
complement(12407..14137)
/note="gene_id:MNJ7.5"
/codon_start=1
/evidence=not_experimental
/product="selenium-binding protein-like"
/protein_id="BAB09072.1"
/db_xref="GI:9758774"
/translation="MLRTVSNFTTRSHVSGTASSNSWTIVPALRFGTGLVRAAV
ELINDGEPRDASPLVHLRVSNGVYVSLCQLHGTYTKGFSVNTSLMSLFYKT
SDSLDAHKVDEMPDPDPIVSNLSVGVQSGRPFQEGCLFLLEHRSDFPNFEST
ALAAACARLHSPGLGACVTHSKVLGKGVVYVGNCLIDMYKCGFMDADVLFQHM
BEKTVSNVATVASCNRKLELGLFWFFHOMPDPDTVTYNELIDAFYKGFDPNNAFQV
LWMPNPNSSWNTLTGTVVNSEKSGEATEFTKMHSSGVRFDEYSISIVLAAVAALA
VPMGSLIHACAKLGLDSRVVASALIDMYSKCMLKHAELMFTWPKRNLIVNWM
ISGYARNDSIEAKLNFQLQERFLKPDRETFNLNLAVCSEVPMVLMGYEMIT
NEYTRKPSVHCCLSRIMGORGVEWQAKQVIFRGVGDVGNRALLGACSRKDLK
RAKTVAAKMIELGADKDEYLIYVNSNLYAYHERVGVQIRKIRRESGVGSSW
IDSRPKCSSY"
join(14542..14786,14885..14950,15046..15162,15242..15464,
15527..15691,15784..15935,16196..16322)
/note="gene_id:MNJ7.6"
/codon_start=1

CDS

/evidence=not_experimental
/product="membrane channel
(tonoplast intrinsic protein)-like"
/protein_id="BAB09071.1"
/db_xref="GI:9758773"
/translation="MVKTEVSGVDSFSVSSLKAYLSSEFIATLLFVAGVGSVAPAK
LTSGLDPAGLVAIAAHAFALFVGVSIANISGHLNPAVTLGLAIGNLTITGF
FYTAQCGSIVACLLVFTNGKSVPTHGVSAGLGAEGVYMEIVTFALVIVYAT
AADPKGSLGTPIATIGFIVGANILAAFPFGSGMNPARSFGPAVVSGLSQIWIY
VGPIVGGALAGLIYGVFISYEAVETREIV"
complement(12407..14137)
/note="gene_id:MNJ7.5"
/codon_start=1
/evidence=not_experimental
/product="selenium-binding protein-like"
/protein_id="BAB09072.1"
/db_xref="GI:9758774"
/translation="MLRTVSNFTTRSHVSGTASSNSWTIVPALRFGTGLVRAAV
ELINDGEPRDASPLVHLRVSNGVYVSLCQLHGTYTKGFSVNTSLMSLFYKT
SDSLDAHKVDEMPDPDPIVSNLSVGVQSGRPFQEGCLFLLEHRSDFPNFEST
ALAAACARLHSPGLGACVTHSKVLGKGVVYVGNCLIDMYKCGFMDADVLFQHM
BEKTVSNVATVASCNRKLELGLFWFFHOMPDPDTVTYNELIDAFYKGFDPNNAFQV
LWMPNPNSSWNTLTGTVVNSEKSGEATEFTKMHSSGVRFDEYSISIVLAAVAALA
VPMGSLIHACAKLGLDSRVVASALIDMYSKCMLKHAELMFTWPKRNLIVNWM
ISGYARNDSIEAKLNFQLQERFLKPDRETFNLNLAVCSEVPMVLMGYEMIT
NEYTRKPSVHCCLSRIMGORGVEWQAKQVIFRGVGDVGNRALLGACSRKDLK
RAKTVAAKMIELGADKDEYLIYVNSNLYAYHERVGVQIRKIRRESGVGSSW
IDSRPKCSSY"
join(14542..14786,14885..14950,15046..15162,15242..15464,
15527..15691,15784..15935,16196..16322)
/note="gene_id:MNJ7.6"
/codon_start=1

alignment_scores:

Quality: 503.00 Length: 190
Ratio: 3.672 Gaps: 5
Percent Similarity: 72.105 Percent Identity: 56.842

alignment_block:

US-09-435-054-2 x AB025628 ..

Align seg 1/1 to: AB025628 from: 1 to: 80117

```

seq_name: gb_pll:AF036684
seq_documentation_block: 826 bp
LOCUS AF036684
DEFINITION Arabidopsis thaliana complete cds.
ACCESSION AF036684
VERSION AF036684.1 GI:3282673
KEYWORDS .
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae;
Magnoliophyta; eudicotyledons;
Brassicales; Brassicaceae;
1 (bases 1 to 826)
Lotan, T., Onto, M., Yee,
Yamauchi, K., Fischer, R.L.,
Arabidopsis LEAF COTYLEDON
development in vegetative
cell 93 (7), 1195-1205
98319234
2 (bases 1 to 826)
Lotan, T., Onto, M.-A., Y
Fischer, R.L., Goldberg,

```


100

1000

```

source
1..197976
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
complement(5668..5772)
/gene="AT4g14420"
/number=1
5668..7235
/gene="AT4g14420"
complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))
/gene="AT4g14420"
/notes="Strong similarity to Nicotiana tabacum ORF able to
induce HR-like lesions, PATX:G1762945
contains EST gb:AI999161.1"
/codon_start=1
/product="elicitor like protein"
/protein_id="CAB78484.1"
/db_xref="GI:7268148"
/translation="MELASELGRALFVSFLLSAWQEFNFGEDGGRSAKSLKPKFNA
FVNHTTHTGOQLPVDVKILVAAALAKGIGLLFVGSLSGLAYLILLHOAVATPIL
YDFNYDVRKEQFLSKFTQSIALGLGLLFTGMKNSRKHGRPKKAPKAKAN"
complement(join(5668..5772,5865..5963,6051..6140,
6238..6352,7168..7235))
/gene="AT4g14420"
complement(5773..5864)
/number=1
complement(5865..5963)
/gene="AT4g14420"
/number=2
complement(5964..6050)
/number=2
complement(6051..6140)
/gene="AT4g14420"
/number=3
complement(6141..6237)
/number=3
complement(6238..6352)
/gene="AT4g14420"
/number=4
complement(6353..7167)
/number=4
complement(7168..7235)
/gene="AT4g14420"
/number=5
complement(8407..9129)
/gene="AT4g14430"
/number=1
complement(9129)
/gene="AT4g14430"
complement(8407..9129)
/gene="AT4g14430"
complement(8407..9129)
/notes="similarity to carnitine racemase, Escherichia coli
contains EST gb:H36845, AI99737.1, T76583"
/codon_start=1
/product="carnitine racemase like protein"
/protein_id="CAB78485.1"
/db_xref="GI:7268149"
/translation="MCTLEKRGDLFLLTLTGDEHRHFDPTTASVLSLLEQAKSQSTR
GSLITITANGKPFNSGFDLAWAQTAGSKTGAANRLHOMVESKPVVAALLDLPMTTIA
ALNGHAAAGLLIALSHDYVFMKRDGVLYMSEVDIGLSPDYPFSAVRAKTGTSAAR
REULLSGKIRGEEAVGLGIVDSAAAYDSEGVVVASVRLGEKLAARKWSGEVYASIRK
SLYPELGLIGLRTVFATPKL"
complement(10443..11159)
/gene="AT4g14440"
/notes="similarity to carnitine racemase, Escherichia coli
contains EST gb:AI994308.1, AA394352, H36791, AA395864,
T76583, H77091, T46265, AA597888, Z17630"
/codon_start=1
/product="carnitine racemase like protein"

```

```

/protein_id="CAB78486.1"
/db_xref="GI:7268150"
/translation="MCTLEKRGDLFLLTLTGDEHRHFDPTTASVLSLLEQAKSQSTR
GSLITITANGKPFNSGFDLAWAQTAGSKTGAANRLHOMVESKPVVAALLDLPMTTIA
ALNGHAAAGLLIALSHDYVFMKRDGVLYMSEVDIGLSPDYPFSAVRAKTGTSAAR
REULLSGKIRGEEAVGLGIVDSAAAYDSEGVVVASVRLGEKLAARKWSGEVYASIRK
SLYPELGLIGLRTVFATPKL"
complement(10443..11159)
/number=1
complement(10443..11159)
/gene="AT4g14440"
complement(10443..11159)
/gene="AT4g14440"
13070..13365
/gene="AT4g14450"
/number=1
13070..14473
/gene="AT4g14450"
join(13070..13365,14296..14473)
/gene="AT4g14450"
/notes="similarity to adenylate cyclase (EC 4.6.1.1) II,
rat
contains EST gb:AF067468, AI994518.1, T44124"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB78487.1"
/db_xref="GI:7268151"
/translation="MASTMORTSSSAASNERQLSQLRAPSLMIKPTSFSSNNVAIP
LISPLAPLSLTSFSDSHVPPQNKTEIPVEEVKTPVFKWQHPASPCYPTPTENR
ASRTSLFDGLGLEGRLRASSYAHDRDNDEALELQDRVSLFKRVCFHLYN"
13366..14295
/gene="AT4g14450"
/number=1
14296..14473
/gene="AT4g14450"
/number=2
17117..17336
17117..22563
/notes="LTR retrotransposon"
17575..22327
/gene="AT4g14460"
join(17575..19093,19145..19476,19552..21256,21338..21472,
21549..22327)
/gene="AT4g14460"
/notes="similarity to similarity to retrovirus-related
polyprotein (retrotransposon Tai-3), Arabidopsis thaliana"
/codon_start=1
/product="retrovirus-related like polyprotein"
/protein_id="CAB78488.1"
/db_xref="GI:7268152"
/translation="MGSVPPANPVTRPPINNGDQIRYPVDQENPYLYLSADHAGL
LLVSRLLTTFDSFHSRRSILMALNVRNKLGFINGITKPPEDHRDFGAWRCNDIVS
TWLMSVDRKKIQGSLLYIATVQGINNLLSRFKODDAPRIFDIQKLSKIQGMDIS
RYIALLLWEEHRYVELPVCCTGCRCCDAALVAKVHQQORSRYTKLSKELNEGFDOT
RRHILMLKPIPTIKEAFNNMTQDERQRNKLTRVDSVAFONTSMINEDENAYVAYN
TVRNPQKPTCHGKVGHTIOCKYKVHCVPGMKTGNTGYTKPNQLHVGPRMPMP
QPRMOPPAQVYTNMOKANVAYVAETGAIYSEGISQAPMNNPYGSPMPHITHGGN
NLSQDFTPOQITEQMSIQFQAQVQVPEAASNNPSPLATVSEHGFALTSTGTTIP
FPSTSLKYENNDLPQNHTLSALQKFLPSDAWIIDSGASSHVCSDLAMFRELKVSQGT
VHTQKLLHNVLDHVPDFKFLMSVSLVKTISCSAHFYVDCCLIQELSOGLMIGRGR
LYHNLYLETENTSPSTPAACLTGSLVNDLHWHORLGHPSVSVLVOKLRLAYIS
HNILASNPFDLHLDIQGPFYESTEGFRYELTVVDDCTRTTWYMLRNKKDSSVFP
EFIKLVSTQFNAKIKAIRSDNAPELGTEIVEKGMHLHFSCATYPOONSVVERKHQH
ILNVAFLFQSNIPMQWSDCVITAVFLINRLPSLNNKSPYELINKPODYSLLK
NFGCLCFVSTNAHERKFTPRACRFLGYPGKYGVLDLSESHVTSRNVVFKKH
VFPKTSLLNKAQDMFENSLPLPAFLHFVETMPLIDEDSLIPTTDSRTADNHASS
SSSALPSITPPSSNETQIDISNAVPIITRSKRTTRAPSLEYHCSILVPSSTPPTD
SSIPHLPELFTASSPKKTPYPISTVVSVDKVTPLCOSYIFAYNETETPKTSQAM
KSEKIRVAVAELOAMELNTWVESLPPDKNVVGCWKVFTIKNPQDGTVERYKARLY
AOGTQBGIDFLDTFSVAKLSAKMMLGLAALITGMTLTQMDVSDAFPLHGDDEEIF
MSLFGYTPPAGTILPPNPVCRLLKSIYGLKQASQWYKRFVAAALVYIDDMTASND
AEVENIKALLRSEFKIKDLGPARFELGLLCKPSSIPMDPTLHLVRDWTPLPNTAY

```



```

RKQSGYGQTKPVFHHKAKTTTKIVLRLOQCCKHFSQRPVKRCKHFEIGDCKGKGT
SLF"
intron complement(745. .828)
exon /number=1
complement(829. .903)
/gene="dl3200c"
intron /number=2
complement(904. .1052)
exon /number=2
complement(1053. .1232)
/gene="dl3200c"
intron /number=3
complement(1233. .1803)
exon /number=3
complement(1804. .1806)
/gene="dl3200c"
exon /number=4
2229. .2837
/gene="dl3205w"
intron /number=1
2229. .7384
/gene="dl3205w"
CDS join(2229. .2837,2914. .3069,3161. .3268,3358. .3468,
3569. .4957,5000. .5212,6340. .6428,6469. .6560,6706. .6749,
7316. .7384)
/gene="dl3205w"
/notes="similarity to kinesin osm-3, Caenorhabditis
elegans
contains EST gb:496806"
/codon_start=1
/product="kinesin like protein"
/protein_id="CAB10212.1"
/db_xref="GI:2244790"
/db_xref="SPTREMBL:O23291"
/translation="NAPTPSSRSNQTQVTLRTPTQKORLNFHSHKTPNPDGSKDPSP
PBPVEVIGRIIDYDPRKEKSPSILQVNTDQTVVRADVGRDFTLDGVSFEQGL
EFYKFIERIKGVKNKCTIMMGTGACKSTMFCCGKEPQIVIRSLDKILGDS
DQGVTFVQVTVLEVNEEYIDLLSTNSNIGIGPKGASTKVRLEVYMKAKNASF
ISGTEAKTSKIVKERRIRKSTICNERSRSHCIIILDVTVGGRLMLVDKAGSE
NIDQAGTFEAKMTAKINQNIALKRVESIANGDSHVPRDSKLTMLQDSDPEDD
KSKILMLCASPDKPMUKTCTLEVGAkakCIVRGSHTPNPKDKYGGDESASAVILGS
RVAWDEFTIKLQSKKKEKEERNEAKOLKKKEVEAALSLTORACATNBEIEK
EYVNRSGSVDETSGFARLRLSLYSDDDPGWVKMSDMDGDPVPKQVWGAHVSHQS
NTISSNLTLPKFSNNLTQMPDRVCLSTVFEVEVEEVEEVEEVEEVEEVEEVEEVEE
PMSLSEGLKGENCFNGADDDKESASRRRLRIQNIETLCGNQRELSHGQDEQANI
ASPDKNQOFFSITNKAEALAEVEAKENNISVDQENGOLDIYVKKWETAADNPKRLIT
TLRVTKDATLADRLKLIETYLGSDNQAFTFLKGLVINLNOQAQAFHEYLPEPCQAV
AKEKSTVQATSLPLCNHAYLATLRPKSSQKSLQSPASPLINPIENMEVTPISK
VTPNHQILKTETERIGEADYINSLSHFLILFPRGESHQHRMLSPHSHOTSPSLSS
FPLLSRADPEGLVDITPPLFEVVLGPTLVLELAVVNDRIAG"
2838. .2913
/gene="dl3205w"
intron /number=1
exon 2914. .3069
/gene="dl3205w"
intron /number=2
3070. .3160
/gene="dl3205w"
exon /number=2
3161. .3268
/gene="dl3205w"
intron /number=3
3269. .3357
/gene="dl3205w"
exon /number=3
3358. .3468
/gene="dl3205w"
intron /number=4
3469. .3568
/gene="dl3205w"
exon /number=4
3569. .4957

```

alignment_scores:

Quality: 410.50

Ratio: 3.310

Percent Similarity: 68.889

Length: 180

Gaps: 4

Percent Identity: 48.333

Quality: 392.50 Length: 191
Ratio: 2.784 Gaps: 5
Percent Similarity: 73.822 Percent identity: 42.932

alignment_block:

US-09-435-054-2 x ZMNFYB ..

Align seg 1/1 to: ZMNFYB from: 1 to: 870

```

6 PheLeuProAlaAlaGlyAlaGluAsnGlySerAlaAlaGly_GlyAlaA 22
||| .....:|||||:.....:|||||:.....:|||||:.....:
4 TTCCGGATAAGCGCAGGAGGAGCTCATGCGGAAGCTCCGGCGACCCCTG 53
.....:|||||:.....:|||||:.....:|||||:.....:|||||:
22 snAsnGlyGlyAlaAlaGlnGlnHisAlaAlaPro.....: 33
.....:|||||:.....:|||||:.....:|||||:.....:|||||:
54 GCGGGCGCGCGGAGCCACGAGAGCGGAGCCCGCCAGGGGAGCGAGGC 103
34 .....AlaIleArgGluGlnAspArgLeuMetProIleAlaAsnVal 48
.....:|||||:.....:|||||:.....:|||||:.....:|||||:
104 GTGGCAGGCTCAGGAGCAGACAGGTTCCTGCCCATCGCAACATCAG 153
48 eArgIleMetArgValLeuProAlaHisAlaLysIleSerAspAspA 65
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
154 TCGCATCATGAGAGGCGATCCGCTCCGAGTTTCATCTCTTCATCACT 203
65 lAlysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThr 81
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
204 CTAAGGAGACCGTGCAGGAGTGCCTCCGAGTTTCATCTCTTCATCACT 253
82 GlyGluAlaAsnGluArgCysGlnArgGluGlnArgLysThrAl 98
.....:|||||:.....:|||||:.....:|||||:.....:|||||:
254 AGCGAAGCAGGTGACAAGTGCAGAGGAGAGCGGAGAGACCATCAATGG 303
98 aGluAspValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValG 115
.....:|||||:.....:|||||:.....:|||||:.....:|||||:
304 CGACGATCTGCTGTGGCCATCGCCACGCTGGGGTTTGAAGACTACATTG 353
115 luProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAla 131
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
354 AACCCCTCAAGGTGACCTACAGAGTACAGAGAGATGGAGGGTGTATAGC 403
132 ArgGlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAsp... 147
AAG .....:|||||:.....:|||||:.....:|||||:.....:
404 AAG.....:TTAATGCTCTAAATCTAGCGATGGCTCGAT 435
148 .....HisHisProHisSerMetSerProAlaAlaM 158
||| ||| ||| ||| |||
436 TAAAGAGATGCTCTTGGTCATGTGGGAGCAAGTACCTCAGCTGCAGAAG 485
158 eLeuLysSerArgGlyProValSer...GlyAlaAlaMetLeuProHis 173
::: .....:|||||:.....:|||||:.....:|||||:.....:
486 GGATGGCCACAGGAGGATACACCAAGGAATGGGTATATATGCAACCT 535
174 HisHisHisHisAspMet 180
.....:|||||:.....:|||||:.....:|||||:.....:
536 CAGTACCATACGGGATATC 556

```

seq_name: gb_p12:ATHAP3A

```

seq_documentation_block: 832 bp mRNA PLN 12-SEP-1997
LOCUS ATHAP3A
DEFINITION Arabidopsis thaliana mRNA for Hap3a transcription factor.
ACCESSION Y13723
VERSION Y13723.1 GI:2398526
KEYWORDS hap3a gene; transcription factor.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 832)
AUTHORS Edwards,D.
TITLE Direct Submission

```

JOURNAL

Submitted (12-JUN-1997) Edwards D., Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK

REFERENCE

2 (bases 1 to 832)

AUTHORS

Edwards,D., Smith,A.G. and Murray,J.A.

TITLE

Isolation and Characterisation of CCAAT box binding proteins from higher plants

JOURNAL

FEATURES

Source

1..832 Location/Qualifiers

/organism="Arabidopsis thaliana"

/db_xref="taxon:3702"

gene

103..528

/gene="hap3a"

CDS

103..528

/gene="hap3a"

/codon_start=1

/product="transcription factor"

/protein_id="CAA74051.1"

/db_xref="GI:2398527"

/translation="MADTSPSPAGDGSGSVREQDRLPIANISIMKKALPPNGK

IGKDAKDTVQECVSEFISFITSEADSKQKPKRTVNGDILLWAMATLGFEDYLEPLK

IYLARYRELEGDNKGSGKSGDGSNRDAGGVSGEEMPSM"

BASE COUNT 252 a 132 c 204 g 244 t

ORIGIN

alignment_scores: Quality: 376.50 Length: 163

Ratio: 3.303 Gaps: 3

Percent Similarity: 69.939 Percent Identity: 46.012

alignment_block:

US-09-435-054-2 x ATHAP3A ..

Align seg 1/1 to: ATHAP3A from: 1 to: 832

```

8 ProAlaAlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnG 24
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
115 CCTTCGAGCCAGCTGGAGATGGCGGAGAAAGCGCGT.....: 153
24 yGlyAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArgL 41
.....:|||||:.....:|||||:.....:|||||:.....:|||||:
154 .....TCCGTTAGGAGCAGGATCGAT 175
41 euMetProIleAlaAsnValIleArgIleMetArgArgValLeuProAla 57
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
176 ACCTTCCTATAGCTAATATCAGCAGGATCATGAAGAAAGCGTTGCCCTCT 225
58 HisAlaLysIleSerAspAlaLysGluThrIleGlnGluCysValSe 74
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
226 AATGGTAAGATTGGAAAGATGCTAAGGATACAGTTCAGGAATCGCTCTC 275
74 rGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArg 91
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
276 TGAGTTTCATCAGCTTCATCAGTACGAGCGCAGTGATAGTGCAAAAG 325
91 luGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArg 107
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
326 AGAAAGAGAAACGTCAATGGTCATGATTTGTGTGGCAATGGCAACA 375
108 LeuGlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTy 124
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
376 TTAGGATTTGAGGATTACCTGGAACCTCTAAAGATATACCTAGCGAGGTA 425
124 rArgGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAla 141
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
426 CAGGAGTTGAGGCGTGATATAAGCGATCAGGA.....: 459
141 laProSerArgGlyGlyAspHisHisProHisSerMetSerProAlaAla 157
|||||:.....:|||||:.....:|||||:.....:|||||:.....:
460 .....AAGAGTCGAGAT.....:GGATCA 477

```

158 MetLeuylsSerArgGlyProValSerGlyAlaMet 170
 478 AATAGAGATGCTGGTGGGTGTTCTCTGTCGAGAAATG 516

seq_name: gb_pl1:AC005309

seq_documentation_block:

LOCUS AC005309 107377 bp DNA PLN 05-APR-2000
 DEFINITION Arabidopsis thaliana chromosome II section 254 of 255 of the
 complete sequence. Sequence from clones T30B22, F17A22, T9J23.
 AC005309 AE02093
 VERSION AC005309.2 GI:6598453
 KEYWORDS HTG.

SOURCE thale cress.

ORGANISM

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons: core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.

REFERENCE

AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
 Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
 Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Renning,C.M.,
 Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
 Tallon,L.J., Gill,J.E., Adams,M.D., Cartera,A.J., Creasy,T.H.,
 Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
 Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
 Venter,J.C.
 Title Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana

JOURNAL

MEDLINE Nature 402 (6763), 761-768 (1999)

PUBMED 20093487

REFERENCE 10617197

AUTHORS 2 (bases 1 to 107377)

TITLE Lin,X.

JOURNAL Direct Submission

Submitted (09-MAR-2000) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

On Dec 17, 1999 this sequence version replaced gi:3738275.
 The sequence and annotation of chromosome 2 were merged from those
 of the individual clones on this chromosome after removing
 overlaps. For detailed information, please see the TIGR web site
 (<http://www.tigr.org/tdb/at/at.html>).

COMMENT

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL
 (<ftp://arthur.epm.ornl.gov/pub/xgrail>), GeneFinder (Phil Green,
 University of Washington), Genscan (Chris Burge,
<http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene
 (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the
 complete sequence against a peptide database and plant EST
 databases at TIGR, and manual curations based on those analyses.
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by two
 or more gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were
 identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are
 numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
 F6P23, F5J6, T1A5, and T1316, the ESSA group for sequencing clone
 F1304, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards
 and Satoshi Tabata for helpful assistance. In addition, we would
 like to thank the TIGR Bioinformatics Department, especially Lixin
 Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
 Peterson, Michael Holmes, and Delwood Richardson for software and
 database support.

This work was supported by the National Science Foundation,
 Department of Energy and the US Department of Agriculture.

Address all correspondence to: atetigr.org.

FEATURES
 source

Location/Qualifiers

1..107377

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

complement(<1..821)

/note="Sequence from clone T30B22"

join(<291..348,504..1024,1120..1549,2112..2233,

2332..>2508)

/gene="At2g47620"

<291..>2508

/gene="At2g47620"

/note="T30B22.7"

join(291..348,504..1024,1120..1549,2112..2233,2332..2508)

/gene="At2g47620"

/codon_start=1

/product="putative SWI/SNF family transcription activator"

/protein_id="AAC63618.2"

/db_xref="GI:6598454"

/translation="MEATDFSAEILYTIPTAQSWFLWDDIHEIRERFAEFTESSI

TRTPKVKYRDFEINKREDTCRLRTFTSVKFLVDVNLQKFLFKWGLINFS

SSLKNDHLLSVDNAKLEQCTPAGIRVTATPNSLRPTVAPLVEERVETGKVPPLTS

YSDVFDLKKPDHVLCAHGCERCDSPFYOHNGKIVNICCKFKNGYNGNNATDFK

LIGNSAAVNTBEEILLLESVLKHGDDEWELISQSVSTKSLDCLISKLILPGEFLM

GSASGRNPSILTEDENETQVDDGDEETETREEDRVNDEPPAKRKRVALISE

SDTLIRASVATLGAQAQKILAQEEREMQLAATVIEQLKQLSKLKLDDLE

SIMDEEKVIEGVKETIIQERSVVLQCAFRSGITKRWDHTYK"

822..107176

/note="Sequence from clone F17A22"

complement(1028..1102)

/rpt_family="(CAAAA)n"

1216..1715

/gene="At2g47620"

/note="molecular marker mi79a"

join(<3826..4057,4175..4305,4398..4535,4614..4712,

4788..>5168)

/gene="At2g47630"

<3826..>5168

/gene="At2g47630"

/note="F17A22.2"

join(3826..4057,4175..4305,4398..4535,4614..4712,

4788..5168)

/gene="At2g47630"

/codon_start=1

/product="putative phospholipase"

/protein_id="AAC63619.1"

/db_xref="GI:3738277"

/translation="MASLVVEKIYSEVDTCKLNKLHVLICGIRLASAGAVFGM

DYEGHSGMSRCYIKKPFANVNDYITSCAQEEDYDKGFLYGESMGAVTLIL

HKKDPFNWAGNLIYAPCKISERKPHPIVNLITRVETIIPKVIPTKVDVDAAFK

DLVKREVRNKLITQDKPRKLTALEMLRTSMNLEDTLHEITPFWFLHGEADVTDP

EVSKALYKASTDKTKLKYGMWHALTSGPCNDVDFADILNMLDLTADPASIT

VTPLRVGNITSVORITVGVSNCHREPKRPFNLLGLNGLRPLVPRSTV"

join(5465..5512,5607..5654,5762..5786,6120..6194,

6385..6489,6572..6844)

/gene="At2g47640"

5465..6844

/gene="At2g47640"

/note="F17A22.3"

join(5650..5654,5762..5786,6120..6194,6385..6489,

6572..6688)

/gene="At2g47640"

/codon_start=1

/product="putative small nuclear ribonucleoprotein D2"

/protein_id="AAC63620.1"

/db_xref="GI:3738278"

/translation="MSKPMBEDTNGKTEEEFNTPGLPSVLMSVKNNTQVLINRNR

KLIGRVRAFRDHCNMVLENREMTVEPKTGKGLKALPVNRDRFISKNFLRGDSVII

18 AlaGlyGlyAlaAsnAsnGlyGlyAlaAla.GlnGlnHisAlaAlaProA 34

18 AlaGlyGlyAlaAsnAsnGlyGlyAlaAla.GlnGlnHisAlaAlaProA 34

overlaps. For detailed information, please visit our website (<http://www.tigar.org/tadb/at/at.html>).

Genes were identified by a combination of three methods: Gene prediction programs including GMAIL (<http://arthur.epm.ornl.gov/pub/xgrail>), GeneFinder (Phil Green, University of Washington), Genscan (Chris Burge, <http://gnomic.stanford.edu/GENSCAN.html>), and NetPlantGene (<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones P6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone P13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES	source	Location/Qualifiers
		1..82403
		/organism="Arabidopsis thaliana"
		/cultivar="Columbia"
		/db_xref="taxon:3702"
		/chromosome="II"
misc_feature		1..82202
		/note="Sequence from clone T10F5"
repeat_region		1197..1306
		/rpt_family="(TAAAA)n"
repeat_region		complement(3226..3308)
		/rpt_family="(TAAAA)n"
repeat_region		7744..7865
		/rpt_family="(TAAAA)n"
gene		complement(7998..11670)
		/gene="At2g13470"
		/note="T10F5.1"
		/pseudo
gene		complement(12098..12827)
		/gene="At2g13480"
		/note="T10F5.2; similar to GB:AAA75253"
		/pseudo
repeat_region		complement(13418..13471)
		/rpt_family="(CAAAA)n"
gene		14215..18083
		/gene="At2g13490"
		/note="T10F5.3"
		/pseudo
repeat_region		complement(18097..18209)
		/rpt_family="(CAAAA)n"
mRNA		join(<18363..19150,19298..>19526)
		/gene="At2g13500"
gene		<18363..>19526
		/note="At2g13500"
		/note="T10F5.4; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"
CDS		join(18363..19150,19298..19526)
		/gene="At2g13500"
		/product="putative Tall-like non-LTR retroelement protein"
		/codon_start=1
		/protein_id="AAD22675.1"
		/db_xref="GI:4558657"
		/translation="MSNVLDIIONELGQDYPTLFTIPNEAYVMEDNRSLIARPLPVRVNEAVINALPRAWGTIAHVHGRILDATYVQVTEIYLLSVORREVMFNMMFVASORWEAAPSLNFVTIDLVQMRGIPILYVFEKTEIAQELGIIISLDHFDATSQTQIAYIRVRVGTIDRLFRQIFESGETALIRFOERLRIYSCFRTHNLNYYPRYQFQSIDREARVDRSLRSSFSOSOMTSSLOVPLHPPRPVSSGDAITSGIMRVFEVGESSRGVQGESHNHNLGESSRKNLVLILORDKESRQHQEQTEQRMGILKPKKR"
	mRNA	complement(join(<20428..20746,20999..>21534))
		/gene="At2g13510"
gene		complement(<20428..>21534)
		/gene="At2g13510"
		/note="T10F5.5; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"
CDS		complement(join(20428..20746,20999..21534))
		/gene="At2g13510"
		/codon_start=1
		/product="putative Tall-like non-LTR retroelement protein"
		/protein_id="AAD22676.1"
		/db_xref="GI:4558658"
		/translation="MADKMDIIONELGPEDPPTLFTIPHEAYVMVAETNRLSIARPLNPRVNLISYIAALPRSGMLTTHVGRVLDATYVQVLFHTMDLVSVORREPLFNRRFVASORWEAPALNFVTIDLVQMRGIPILYVFEETALEIAQELGIIISLDHFDATSQTQIAYIRVRVQIGITERLVRPAPPLNPEELAGASPPFSSRNAAFQYVWVPSQGVGRHVFVPSDINPSSGAITSGVRRVYEVGESFRPCPDGESSRNNRNIGESSRKNLVLQSDNEKAN"
	repeat_region	21554..21667
		/rpt_family="(TAAAA)n"
gene		complement(21749..25813)
		/gene="At2g13520"
		/note="T10F5.6"
		/pseudo
gene		complement(25902..26957)
		/gene="At2g13530"
		/note="T10F5.7"
		/pseudo
mRNA		join(<27480..27563,27648..27707,27807..27863,28349..28420,28529..28591,28765..28941,29126..29254,29478..29544,29724..29923,30139..30209,30672..30756,30903..30971,31063..31181,31489..31627,31829..32329,32420..32778,33154..33372)
		/gene="At2g13540"
		<27480..33372
		/gene="At2g13540"
		/note="T10F5.8"
		join(27480..27563,27648..27707,27807..27863,28349..28420,28529..28591,28765..28941,29126..29254,29478..29544,29724..29923,30139..30209,30672..30756,30903..30971,31063..31181,31489..31627,31829..32329,32420..32778)
		/gene="At2g13540"
		/codon_start=1
		/product="putative cap-binding protein"
		/protein_id="AAD22677.1"
		/db_xref="GI:4558659"
		/translation="MSNMKTLRLRIGKPEYGTSSDYKDIETFCGIVIRREIERSGDQVLPFLIQCAQLPKPIPLYGTGLIGLLENEDFVKLVESVHANFOALDSGNCNRI RILLRFMTSLICKSVIOPASLIVVFETLLSAAATVDEKGNPSQPOADPYVICILS LIPWGSSELAFOVPDEIRVLVGIQAYLSIRKNSSTSLGNFFHNGEPESSEIAKDFVE DLDRIOQLASNGHKLSESVPRPHLSFEAQLVAGKFHELRIKCMQEQSPSSDISRAIS GKQKDALTRYPYLRKLNIFPPANKMEDVQPIDRVFVEEYLLDVLFLYNGCRKECASV MANLPVTFVRYLMAETLFSQILLPQPPTKTYTLVIMDLCKALPGAPVAVVAGV RALFEKISDLDMESRTRILILWFSHLSNFQIPWEEWAFVLDLPKAPKRVFOEITL QREVLRSYWDKIKOSIENATALELLPKAGPNPMYSLEEGKEKTEQOOLSAELSRKV KEKOTARDMLVMEETIYPVHGFEVTLTVQVLLDGLSKSFHLTVLRYGOVFSK LCPNDQVQLLSQVSYIWNKNVOMTAVADRMGMVRLVSNQALVRWFSPENVDQPH VSDQFWEIQLGNALNKTYNRISDLRKSINITKNVLYVAEKASANARVLEAAESKLSIV EGEVIGENPANKMKRLKSTVEKTGAELSRESLEAKELINRLALSETELYISLMQH MILLFPFTY"

```
repeat_region complement(34637..34727)
mRNA /rpt_family="(TA)n"
gene /complement(38932..40224)
     /gene="At2g13550"
     /complement(38932..40224)
     /gene="At2g13550"
     /note="T10F5_9"
CDS complement(39231..39776)
     /gene="At2g13550"
     /note="unknown protein"
     /codon_start=1
     /protein_id="AAD22678.1"
     /db_xref="GI:458660"
     /translation="MTSQIRQTPLRMHLTQFPNTRGRQPKSPNSHHDPDESPS
POQRHHTREPIRTHQKLTSSSTANTFTGDKRKHIVTNDIADTKTRTS
RSGNQSPQPHIQKLSQSLKTAANNIENLRNRRREIGRPTQHQHPYNNHITKT
RRKSTDEIKORTRPSVLR"
mRNA join(<40516..40686,40766..40862,40943..41061,41171..41338,
41499..41627,41718..41778,42210..42313,42390..42473,
alignment_scores:
  Quality: 354.50 Length: 209
  Ratio: 2.686 Gaps: 6
Percent Similarity: 63.158 Percent Identity: 37.799
alignment_block:
US-09-435-054-2 x AC007063/rev ..
Align seg 1/1 to reverse of: AC007063 from: 1 to: 82403
8 ProAlaAlaGlyAlaGluAsn...GlySerAlaAlaGlyGlyAlaAsnAs 23
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
46880 CTGGAGTAGCTGAACAATAATCCAGGAAGCCCTTCTTCAAGACCAACAA 46831
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
23 nGlyGlyAlaAlaGlnGlnHisAlaProAlaIleArgGluGlnAspA 40
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46830 CACACACACACACAC.....AAAGACAGGACC 46802
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
40 rGluMetProIleAlaAsnValIleArgIleMetArgValLeuPro 56
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46801 GGTTCCTCCATTCGGAATTCGGAAGGATCATGAAAGAGTTCTTCCC 46752
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57 AlaHisAlaIleSerAspAlaGlyGluThrIleGlnGluCysVa 73
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46751 GGTAACGGTAAGATCTCAAAAGACGCTAAAGAAACCGTTCAAGAATGT 46702
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 lSerGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnA 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46701 CTCGGAGTTCATTAGTTTCGTCTGCTGCTGCTGCTGCTGCTGCTGCT 46652
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
90 rGluGlnArgIleThrIleThrAlaGluAspValLeuTrpAlaMetSer 106
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46651 GAGAAAGAGAGAGAGACCATCAATGGAGATGATATCATTTGGGCTATCACA 46602
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
107 ArgLeuGlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisAr 123
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46601 ACTCTCGGTTTCGAGAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 46552
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123 gTyrArgGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyA 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46551 ATATAGACACCGAAGGAGAGAAAGTTAAC..... 46521
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140 lAlaProSer.....ArgGly 145
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46520 ..AGCCAAACAAACAAACAAACAAACAAACAAACAGCAGATTCAACAA 46473
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 GlyAspHisHisProHisSerMetSerProAlaAlaMetLeuLysSerAr 162
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46472 CAGAATCATCAATAATTATCAGTTTCAAGAACCAAGAC.....CAAAACAA 46429
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
162 gGlyProValSerGlyAlaAlaMetLeuProHisHisHisHisHisA 179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
46428 TAACACATGTCATGTACTAGTTACATCTCTCATCATCAT..... 46389
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```

|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
243 AATGAAATGCTGTACCACAGCAGGAAATAATTCACAAAGACGCAAGG 292

67 luThrIleGlnGluCysValSerGluThrIleSerPheIleThrGlyGlu 83
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
293 AGTGTGTACAGGAATGCTAAGCGAATTTATAGCTTCAATACATCTGAA 342

84 AlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAs 100
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
343 GCAAGTGGAGGATGCCACAGGAGGAGGAGCAATGAAAGGGAAGGC 392

100 pValLeuTrpAlaMetSerArgLeuGluPheAspAspTyrValGluProL 117
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
393 TATTCATTGCTATGTCGAGACTAGGCTTTGATAGTATGATAGCCAT 442

117 euGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArgGly 133
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
443 TAAAGTTGTATCTCTCAGAAATTCAGAGAGGCAATGAAAGGGAAGGC 492

134 ValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAsp.....HI 148
:|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
493 ATTGGTGTACAGTTACACAGAGAGATGCTGGCGGAGAGACCTTCGAGA 542

148 sHisProHis...SerMetSerProAlaAlaMetLeuLysSerArgGlyP 164
:|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
543 AGAACCTTTCACTAGCCAAATACACAGCAGGTTTAATAACACAGATGGAC 592

164 roValSerGlyAlaAlaMet..LeuProHisHis 174
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
593 AGCAACAGATGTTATGGTTTATACCATCAT 625

```

seq_name: gb_ov:PMNFB

```

seq_documentation_block: 699 bp mRNA VRT 09-FEB-1999
LOCUS PMNFB
DEFINITION P.marinus mRNA for CAAT-box DNA binding protein subunit B (NF-YB).
ACCESSION X59712
VERSION 1 (bases 1 to 699)
KEYWORDS CAAT-box DNA binding protein.
SOURCE sea lamprey.
ORGANISM Petromyzon marinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
Petromyzontiformes; Petromyzontidae; Petromyzon.
1 (bases 1 to 699)
Direct Submission
Submitted (21-JAN-1992) C. Benoist, L.G.M.E., Dept of Immunology,
11, Rue Humann, Strassbourg 67000, FRANCE
Li.X.Y., Mantovani,R., Hooft van Huijsdijnen,R., Andre,I.,
Benoist,C. and Mathis,D.
Evolutionary variation of the CCAAT-binding transcription factor
NF-Y
Nucleic Acids Res. 20 (5), 1087-1091 (1992)
JOURNAL 92195809
MEDLINE
REMARK Erratum:[published erratum appears in Nucleic Acids Res 1992 Apr
11;20(7):1841]]

```

```

FEATURES
source
1..699
/organism="Petromyzon marinus"
/db_xref="taxon:7757"
40..669
/codon_start=1
/product="CAAT-box DNA binding protein subunit B (NF-YB)"
/protein_id="CA4232.1"
/db_xref="GI:64218"
/db_xref="SWISS-PROT:P25210"
/translaton="MDAGSTTDASOLGITDYIGGGHVYVLOSSDGDAGSLAGSDHD
ESGSKDPYREQDIYLPANVARIMKTSIPSSGKIAKDAKEQVCEVSEIFSIPISEA
SERCHQEKRTINGEDILFAMSLGDFSVPEPLKQYKYSKMGKEGINATVVTIT
DAIPELTESFSGPLATSIITADGQQQNVMTTATYIQIGVQPIQFT"
BASE COUNT 186 a 179 c 205 g 129 t
ORIGIN

```

```

alignment_scores:
Quality: 325.50 Length: 124
Ratio: 3.356 Gaps: 1
Percent Similarity: 78.226 Percent Identity: 53.226

alignment_block:
US-09-435-054-2 x PMNFB
..
Align seg 1/1 to: PMNFB from: 1 to: 699

11 GlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGlyAlaAl 27
|| .....:|||||.....:|||||.....:|||||.....:|||||.....:
133 GGTGATGCTGAGGTTAGCTGCGAGTGTGTACGATGAGAGAGCTGTGG 182

27 acGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArgLeuMetProI 44
:|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
183 ATCGAAA.....GACCCGTATCGCGAGCAGACATCTACCTGCCGA 223

44 leAlaAsnValIleArgIleMetArgArgValLeuProAlaHisAlaLys 60
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
224 TAGCTAACGTGCCCGGATCATGAAGACTCATCCATCCCTCTCTGGGAG 273

61 IleSerAspAlaLysGluThrIleGlnGluCysValSerGluTyrIle 77
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
274 ATTGCAAGGAGCGCAAGGAGTGTGTGCGAGGAGTGGTGAGCGAATTCAT 323

77 eSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluGlnArgL 94
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
324 CAGCTTCATCATCGTGGAGCGGAGCGAACGCTGCCACCGAGGAAACGCA 373

94 ysThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeuGlyPhe 110
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
374 AGACCATCAATGGGAGGAGACATCTCTTCGCGCATGTCCACCTCTGGCTC 423

111 AspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrArgGluPh 127
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
424 GACAGTACGTGCGCGCTCAAGCAGTATCTTCAAAAGTACCGTGAGTC 473

127 eGluGlyAspAlaArgGlyVal 134
|||||.....:|||||.....:|||||.....:|||||.....:|||||.....:
474 AATGAAGGCTGAGAAAGGAATC 495

```

seq_name: gb_ro2:RATCBFALL

```

seq_documentation_block: 538 bp mRNA ROD 27-APR-1993
LOCUS RATCBFALL
DEFINITION Rat CCAAT binding transcription factor-B subunit (CBF-B1) mRNA,
complete cds.
ACCESSION M60617.1 GI:203354
VERSION 1
KEYWORDS CCAAT binding transcription factor-B subunit.
SOURCE Rat (Sprague-Dawley) liver, cdna to mRNA.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 538)
AUTHORS Vuorio,T., Maiti,S.N. and de Crombrughe,B.
TITLE Purification and molecular cloning of the A chain of a rat
heteromeric CCAAT-binding protein: Sequence identity with the yeast
HAP3 transcription factor
J. Biol. Chem. 265, 22480-22486 (1990)
FEATURES
91093096
Location/Qualifiers
1..538
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
/tissue_lib="lambda-ZAP"
1..447
gene

```

```

JOURNAL
MEDLINE
FEATURES
Source
1..538
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/tissue_type="liver"
/tissue_lib="lambda-ZAP"
1..447
gene

```

```

CDS
/contig="CBF-All"
1..447
/contig="CBF-All"
/codon_start=1
/product="CCAAT binding transcription factor-B subunit"
/protein_id="AAA40888.1"
/db_xref="GI:203355"
/translation="MNDHEDTNGSKSFREDDIYLPITANVARIMKNAIPOTGKIADAK
KECVQECVSEFISITSEASERCHQERKKTINGEDILFAMSTLGFDSYVEPLKLYLOK
FREAKGKGGIGGAVSATDGLSELTETAEFSKLTAAAGFSPWAVC"
BASE COUNT 158 a 106 c 152 g 122 t
ORIGIN
```

```

alignment_scores:
  Quality: 324.50      Length: 113
  Ratio: 3.527        Gaps: 1
  Percent Similarity: 81.416  Percent Identity: 56.637

alignment_block:
US-09-435-054-2 x RATCBFAll ..
Align seg 1/1 to: RATCBFAll from: 1 to: 538

34 AlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgI1 50
   ::::|||||
37 AGTTTCAGAGACAAGATATTATCTCCCATTCGCAATGCAATGTGGCTAGGAT 86
   |||:|||||
50 eMetArgValLeuProAlaHisAlaLysIleSerAspAlaLysG 67
   |||:|||||
87 AATGAAAATGCCATACCTCAACAGGAAAGATTGCAAGGATGCCAAAG 136
   |||:|||||
67 luThrIleGlnGlyCysValSerGluTyrIleSerPheIleThrGlyGlu 83
   |||:|||||
137 AATCCGTTACAGAGTGTGTCAGTGTGATTTATAGCTTCATACATCTGAA 186
   |||:|||||
84 AlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAs 100
   |||:|||||
187 GCAAGTGAGCGCTCTCACCAGGAGAGCGCAAGACCATCAACGGGGAGGA 236
   |||:|||||
100 pValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValGluProL 117
   |||:|||||
237 CATTCGTGCGCATGTCACACTCTCGGCTTCGACAGCTACGTGGAGCCTC 286
   |||:|||||
117 euGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArgGly 133
   |||:|||||
287 TGAACACTGCTACCTCCAGAAGTTCAGAGAGGCCCATGAAGGAGAGAGGCC 336
   |||:|||||
134 ValGlyLeuValProGlyAlaAlaProSerArgGlyGly 146
   |||:|||||
337 ATTGGT.....GGGGCGGTGTCTGTCTACAGATGGA 366
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 25, 2001, 10:47:45 ; Search time 1233.41 Seconds
(without alignments)
14710.169 Million cell updates/sec

Title: US-09-435-054-1
Perfect score: 1173
Sequence: 1 ccacgcgtccgccaccacac.....tcaaaaaaaaaaaaaaaaaa 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues
Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_bal:*
- 2: gb_ba2:*
- 3: gb_ba3:*
- 4: gb_in1:*
- 5: gb_in2:*
- 6: gb_in3:*
- 7: gb_om:*
- 8: gb_ov:*
- 9: gb_pat1:*
- 10: gb_pat2:*
- 11: gb_ph:*
- 12: gb_pl1:*
- 13: gb_pl2:*
- 14: gb_pl3:*
- 15: gb_pl4:*
- 16: em_bal:*
- 17: em_ba2:*
- 18: em_fun:*
- 19: em_htgo_hum:*
- 20: em_htgo_inv:*
- 21: em_htgo_rod:*
- 22: em_htg_hum1:*
- 23: em_htg_hum2:*
- 24: em_htg_hum3:*
- 25: em_htg_hum4:*
- 26: em_htg_hum5:*
- 27: em_htg_hum6:*
- 28: em_htg_hum7:*
- 29: em_htg_hum8:*
- 30: em_htg_inv1:*
- 31: em_htg_inv2:*
- 32: em_htg_other:*
- 33: em_htg_rod:*
- 34: em_hum1:*
- 35: em_hum2:*
- 36: em_hum3:*
- 37: em_hum4:*
- 38: em_hum5:*
- 39: em_hum6:*
- 40: em_hum7:*
- 41: em_in:*
- 42: em_om:*
- 43: em_or:*

44: em_ov:*

45: em_pat:*

46: em_ph:*

47: em_pl:*

48: em_ro:*

49: em_sts:*

50: em_sy:*

51: em_un:*

52: em_vi:*

53: gb_sts1:*

54: gb_sts2:*

55: gb_sts3:*

56: gb_sy:*

57: gb_un:*

58: gb_vil:*

59: gb_vil2:*

60: gb_htg1:*

61: gb_htg2:*

62: gb_htg3:*

63: gb_htg4:*

64: gb_htg5:*

65: gb_htg6:*

66: gb_htg7:*

67: gb_htg8:*

68: gb_htg9:*

69: gb_htg10:*

70: gb_htg11:*

71: gb_htg12:*

72: gb_htg13:*

73: gb_htg14:*

74: gb_htg15:*

75: gb_htg16:*

76: gb_htg17:*

77: gb_htg18:*

78: gb_htg19:*

79: gb_htg20:*

80: gb_htg21:*

81: gb_htg22:*

82: gb_htg23:*

83: gb_htg24:*

84: gb_htg25:*

85: gb_pr1:*

86: gb_pr2:*

87: gb_pr3:*

88: gb_pr4:*

89: gb_pr5:*

90: gb_pr6:*

91: gb_pr7:*

92: gb_pr8:*

93: gb_pr9:*

94: gb_rol:*

95: gb_rol2:*

96: gb_in4:*

97: gb_pr10:*

98: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	199.8	17.0	141808	83	AP003266
2	199.8	17.0	179686	83	AP003246
3	198.6	16.9	80117	12	AB025628
4	179	15.3	826	12	AF036684
c 5	177.4	15.1	82875	12	AC013482
6	158.2	13.5	870	15	ZMNFYB
7	142.2	12.1	874	13	ATHAP3B
8	138.2	11.8	699	8	PMNFYB
					AP003266 Oryza sat
					AP003246 Oryza sat
					AB025628 Arabidops
					AF036684 Arabidops
					AC013482 Genomic s
					X59714 Z.mays mrna
					Y13724 Arabidopsis
					X59712 P.marinus m

TITLE Direct Submission
JOURNAL Submitted (19-FEB-2001) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@abrr.affrc.go.jp, URL:http://rtp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT NOTE: It currently consists of 1 contigs. Gaps between the contigs
 are represented as runs of N. The order of the pieces is believed
 to be correct as given, however the sizes of the gaps between them
 are based on estimates that have provided by the submitter. This
 sequence will be replaced by the finished sequence as soon as it is
 available and the accession number will be preserved.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
FEATURES Location/Qualifiers
 source
 1..179686
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0423A12"
BASE COUNT 51983 a 38309 c 38396 g 50748 t 250 others
ORIGIN
 Query Match 17.0% Score 199.8; DB 83; Length 179686;
 Best Local Similarity 65.7%; Pred.No. 3.2e-20;
 Matches 291; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
 Qy 136 atggcggtgtctcagcagcagtcggcggtgctccgagcagcagcgtgctgac 195
 Db 28008 AGGGGGAGCGCGGATCAGCGCGCGGAGATCATCAAGGAGCAGGAGTCTGTC 28067
 Qy 196 cgatcgcaacgtgacgtatcagcggtgctgctgctgctgctgctgctgctg 255
 Db 28068 CAATCGGAGCGTGGGCGGAATCATGAAGCAGATCTCTCCCGCCCAAGCGCAAGATCTCCA 28127
 Qy 256 acgacgcaaggagacgacgacgagtgctgctgctgctgctgctgctgctgctg 315
 Db 28128 AGGAGGCAAGGAGCAGCAGGAGTCTGCTCGGAGTTCATCAGCTTCACCGCGG 28187
 Qy 316 aggcacacgagcgtgctgagcgtgagcgtgagcgtgagcgtgagcgtgagcgt 375
 Db 28188 AGGCTTCGACAGTCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 28247
 Qy 376 gggcctgagcgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 435
 Db 28248 GGGCTTCGGCGCGCTCGGCTTCGACGACTGCTGACCGCCCTCCGCGCGCG 28307
 Qy 436 gctaccgaggttcgagggcgagcgtgctgctgctgctgctgctgctgctgct 495
 Db 28308 AGTACCGAGGCTCGAGGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 28367
 Qy 496 cgcgcgcgcgac 555
 Db 28368 GCBCG 28427
 Qy 556 ggcagctctcggagcgccatg 578
 Db 28428 GTCAATTCATGTTCAAGGCCATG 28450
RESULT 3
LOCUS AB025628
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, p1 clone:MNJ7.
ACCESSION AB025628 BA000015
VERSION AB025628.1 GI:4589434
KEYWORDS
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui p1
 clone:MNJ7.

ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsi.
 1 (sites)
 Kaneko,T., Katoh,T., Asamizu,E., Sato,S., Nakamura,Y., Kotani,H.
 and Tabata,S.
 Structural analysis of Arabidopsis thaliana chromosome 5. XI
 Unpublished
 2 (bases 1 to 80117)
 Nakamura,Y.
 Direct Submission
 Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research; 1532-3, Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
 Tel:81-438-52-3935, Fax:81-438-52-3934)
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=mnj7
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://grmln1.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is MQL5 and the 3' clone is MGL1.
FEATURES Location/Qualifiers
 1..80117
 /organism="Arabidopsis thaliana"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MNJ7"
 /clone_lib="Mitsui p1"
 complement(12..108)
 /note="CDS is reported in Acc# AB018117"
 gene_id:MQL5.29
 /number=2
 /evidence=not_experimental
 /product="DNA-binding protein-like"
 complement(375..540)
 /note="CDS is reported in Acc# AB018117"
 gene_id:MQL5.29
 /number=1
 /evidence=not_experimental
 /product="DNA-binding protein-like"
 join(1723..1976,2116..2266,2344..2417,2508..2667,
 2834..2932,3028..3114,3222..3365)
 /note="gene_id:MNJ7.1"
 /codon_start=1
 /pseudo
 /evidence=not_experimental
 /product="formyltetrahydrofolate deformylase-like"
 complement(join(3514..3755,3836..3893,3983..4032,
 4110..4289,4380..4773,4852..5148))
 /note="gene_id:MNJ7.2"
 join(171442
 p1r1|G71442
 strong similarity to unknown protein"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="BAB09069.1"
 /db_xref="GI:9758771"
 /translation="MEGGFYSDWDSLSLFGSENPHELEGNVRSSEIVSIPOQ"

exon

exon

CDS

CDS

COMMENT

[illegible]

RESULT	6	
ZMNFYB		
LOCUS	870 bp	mRNA
DEFINITION	2.mays mRNA for CAAT-box DNA binding protein subunit B (NF-YB).	PLN 09-FEB-1999
ACCESSION	X59714	
VERSION	1	GI:22379
KEYWORDS	CAAT-box DNA binding protein.	
SOURCE	zea mays.	
ORGANISM	zea mays	
REFERENCE	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.	
AUTHORS	1 (bases 1 to 870)	
TITLE	Benoist,C.	
JOURNAL	Direct Submission	
REFERENCE	Submitted (21-JAN-1992) C. Benoist, L.G.M.E., Dept of Immunology, 11, Rue Humann, Strassbourg 67000, FRANCE	
AUTHORS	2 (bases 1 to 870)	
TITLE	Li,X.Y., Mantovani,R., Hooft van Huijsduijnen,R., Andre,I., Benoist,C. and Mathis,D.	
JOURNAL	Evolutionary variation of the CCAAT-binding transcription factor NF-Y	
MEDLINE	Nucleic Acids Res. 20 (5), 1087-1091 (1992)	
REMARK	92195809	
FEATURES	Erratum: [[published erratum appears in Nucleic Acids Res 1992 Apr 11;20(7):1841]]	
source	Location/Qualifiers	
	1..870	
	/organism="Zea mays"	
	/db_xref="taxon:4577"	
	29..565	
	/codon_start=1	
	/product="CAAT-box DNA binding protein subunit B (NF-YB)"	
	/protein_id="CA4234.1"	
	/db_xref="GI:22380"	
	/db_xref="SWISS-PROT:P25209"	
	/translation="MAEPASPGGGGSHSGSPRGGGGGVSREQDRFLPTIANISRI MKKAIIPANKIKADAKETVOECVSSEFISITSEADSKCQREKRTINGDLLWAMATL FDFDYTEPLKVLQGRYREMEGDSKLTAKSGDSGITKDALGHVAGSSAABGMSQQAQ NCQGMGYMPQYHNGDISN"	
BASE COUNT	224 a 189 c 258 g 199 t	
ORIGIN		

Query Match	13.5%;	Score 158.2;	DB 15;	Length 870;																																																																																															
Best Local Similarity	68.2%;	Pred. No. 2.7e-13;																																																																																																	
Matches 234;	Conservative 0;	Mismatches 108;	Indels 1;	Gaps																																																																																															
QY	115	cggcgccgcccgcgcacaaatcggcgccgtctcagcagcatgccccgccgagtc	174																																																																																																
					QY	175	gcgagcaggaccggcgtgatccgatcgcgaactgataccgcatacatcgcgcgctgctgc	234							D	117	GGGAGCAGGCAGATTCTTCGCCCATCCGCCAATCATCTGCATCATGAAGAAGGCCATCC	176							QY	235	cggcgacgcacaagtatctcggacgcgcgaaggagacgatccaggagtgctgtcgcgagt	294							D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399	
QY	175	gcgagcaggaccggcgtgatccgatcgcgaactgataccgcatacatcgcgcgctgctgc	234							D	117	GGGAGCAGGCAGATTCTTCGCCCATCCGCCAATCATCTGCATCATGAAGAAGGCCATCC	176							QY	235	cggcgacgcacaagtatctcggacgcgcgaaggagacgatccaggagtgctgtcgcgagt	294							D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399						
					D	117	GGGAGCAGGCAGATTCTTCGCCCATCCGCCAATCATCTGCATCATGAAGAAGGCCATCC	176							QY	235	cggcgacgcacaagtatctcggacgcgcgaaggagacgatccaggagtgctgtcgcgagt	294							D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399											
D	117	GGGAGCAGGCAGATTCTTCGCCCATCCGCCAATCATCTGCATCATGAAGAAGGCCATCC	176							QY	235	cggcgacgcacaagtatctcggacgcgcgaaggagacgatccaggagtgctgtcgcgagt	294							D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																
					QY	235	cggcgacgcacaagtatctcggacgcgcgaaggagacgatccaggagtgctgtcgcgagt	294							D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																					
QY	235	cggcgacgcacaagtatctcggacgcgcgaaggagacgatccaggagtgctgtcgcgagt	294							D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																										
					D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																															
D	177	CGGTAAACGGGAGGATCGCCAAAGGAGCCTAAGGAGACCGTGTCAGGAGTGCCTCCGAGT	236							QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																				
					QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																									
QY	295	acatcaattcatacgcggggaggccaacgagcggctcgaagcgggagcgcgaagacca	354							D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																														
					D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																			
D	237	TCAATCTCTTATCATCTAGCGGACGAGTAGTGCACAAGTGCCACAGGGGAGAGCGGAACACCA	296							QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																								
					QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																													
QY	355	tcaocgcggaggacgtgctgtggccatgagcgccctcggttctcgacgactagctcgagc	414							D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																		
					D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																							
D	297	TCAATGCGAGCATCTGCTGTGGCCATGGCCACGCTGGGGTTTGAAGACTACATTGTAAC	356							QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																												
					QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																																	
QY	415	cgtcggcgacctaacccgcctaocgcgagttctgaggcga	457							D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																																						
					D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																																											
D	357	CCCTCAAAGGTGTACCTACAGAAGTACAGAGAGATGGAGGGTGA	399																																																																																																

RESULT	7	
ATHAP3B	874 bp	mrna
LOCUS	Arabidopsis thaliana	mRNA for Hap3b transcription factor.
DEFINITION	Y13724	
ACCESSION	Y13724.1	GI:2398528
VERSION	hap3b gene; transcription factor.	
SOURCE	thale cress.	
ORGANISM	Arabidopsis thaliana	
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Brassicales; Brassicaceae; Arabidopsids.	
REFERENCE	1	(bases 1 to 874)
AUTHORS	Edwards,D.	
TITLE	Direct Submission	
JOURNAL	Submitted (12-JUN-1997) Edwards D., Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK	
REFERENCE	2	(bases 1 to 874)
AUTHORS	Edwards,D., Smith,A.G. and Murray,J.A.	
TITLE	Isolation and characterisation of CCAAT box binding proteins from higher plants	
JOURNAL	Unpublished	
FEATURES	Location/Qualifiers	
Source	1..874	
	/organism="Arabidopsis thaliana"	
gene	/db_xref="taxon:3702"	
	1..565	
	/gene="hap3b"	
CDS	<1..565	
	/gene="hap3b"	
	/codon_start=2	
	/product="transcription factor"	
	/protein_id="CAA74052.1"	
	/db_xref="GI:2398529"	
	/db_xref="SPTREMBL:O23634"	
	/translation="NRDSGGGQNGNQSSLSPRODFRFLPIANVSRIMKKAALPA NAKISKDAKTEMOECVSEFISRTVEASDKCKQKERRKTINDGDLWMTTIGPDEVYE PLKVIYLRFRTEIGERTGLRPTGTGVEGHRDAVGGGGGYGGGGMOYHOHHQFL HQONHYIGATGGDSGGGAASGRTRT"	
BASE COUNT	263 a	155 c 244 g 212 t
ORIGIN		

Query Match 12.18; Score 142.2; DB 13; Length 874;

Query Match	11.8%;	Score 138.2;	DB 8;	Length 699;
Best Local Similarity	68.5%;	Pred. No. 1.9e-10;		
Matches 191;	Conservative	0;	Mismatches 88;	Indels 0;
Gaps	0;			
Qy	169	cgatccgcagcaggaccgctgatccgcgatcgcgaacgtatccgcattcatcgcgcgcg	228	
		/gene="AT4g14420"		
		complement(5773..5864)		
		intron		
		gene		
		6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		
		complement(join(5668..5772,5865..5963,6051..6140,		
		5772,5865..5963,6051..6140,		
		5963,6051..6140,		
		6140,6238..6352,7168..7235))		

AUTHORS **TITLE** **JOURNAL**

COMMENT

Lin, X.
Direct Submission
Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Dec 17, 1999 this sequence version replaced gi:3738275.
The sequence and annotation of chromosome 2 were merged from those
of the individual clones on this chromosome after removing
overlaps. For detailed information, please see the TIGR web site
(<http://www.tigr.org/tdb/at/at.html>).

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL
(<http://arthur.epm.ornl.gov/pub/xgrail/>), Genefinder (Phil Green,
University of Washington), Genscan (Chris Burge,
<http://gnomic.stanford.edu/GENSCANW.html>), and NetPlantGene
(<http://www.cbs.dtu.dk/services/NetGene2/>), searches of the
complete sequence against a peptide database and plant EST
databases at TIGR, and manual curations based on those analyses.
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by two
or more gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats were
identified by RepeatMasker (Arian Smit,
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). Genes are
numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones
F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone
F13D4, and Scott Jackson, Jining Jiang, Klaus Meyer, Eric Richards
and Satoshi Tabata for helpful assistance. In addition, we would
like to thank the TIGR Bioinformatics Department, especially Lixin
Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy
Peterson, Michael Holmes, and Delwood Richardson for software and
database support.

This work was supported by the National Science Foundation,
Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

Location/Qualifiers
1. 107377

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

complement(<1..821)

/note="Sequence from clone T30B22"

join(<291..348,504..1024,1120..1549,2112..2233,

2332..>2508)

/gene="At2g47620"

<291..>2508

/gene="At2g47620"

/note="T30B22.7"

join(291..348,504..1024,1120..1549,2112..2233,2508)

/gene="At2g47620"

/codon_start=1

/product="putative SWI/SNF family transcription activator"

/protein_id="AAC63618.2"

/db_xref="GI:6598454"

/translation="MEATDPSAEIYLTIPAOSSWFLWDDIHIERREFAEPTESSI

TRPKVYKEDRIINKFREDTCRRITFTSVRFKLVGDVNLQKVFLEKGLNFS

SSLKNDHLLSVKAEHQCTPAGIRTPNLSRIPAPPLVEERVTGKVPPTLS

YSDVFDLKPDLVCAHGCRCDSPTFOHNGKIVNICKEKFNKNGYENGTADDFK

LIGNSAAVTEEEILLLESLVKGHDWELISQSVSTRKSDICLIELPFGFEL

LGASGRNLNPTEDTEOVQTDGHEHETETREKEDRNDDEPPAKRKRVLTSE

GDTLIRASVATAGAAAOAKTLAQDEEREMEQLAATVLEOQLKLSKLFLLDLE

SIMDEEKVTEGKETIIQERVSVLQAFRSGITRKWDHTYVK"

822..107176

misc_feature

repeat_region

misc_feature

mRNA

gene

CDS

/note="Sequence from clone F17A22"

complement(1028..1102)

/rpt_family="(CAAAA)n"

1216..1715

/gene="At2g47620"

/note="molecular marker mi79a"

join(<3826..4057,4175..4305,4398..4535,4614..4712,

4788..>5168)

/gene="At2g47630"

<3826..>5168

/gene="At2g47630"

/note="F17A22.2"

join(3826..4057,4175..4305,4398..4535,4614..4712,

4788..5168)

/gene="At2g47630"

/codon_start=1

/product="putative phospholipase"

/protein_id="AAC63619.1"

/db_xref="GI:3738277"

/translation="MASLVVEKTIYSEVDTCLKNLKHVLILECGIRLASAGYAVFGM

DYEGHGRSMGRSRYIKKFNANVYNDYDYSICAEYEMDKGRFLYGSMGAVTLLL

HKKPLFWNGALLVAPMCKISSEKPAFIVINLTVREIEIPKPKIVPTKVIDRAFK

DLVKEEVNKKLIYQDKPRKTALEMLRTSNLLEDTLHEITMPEFVJHGSDVTPD

YSKALYKASTRDKTLKYPGMHALTSGEPCNVLDVFADIIINLDRTPASLT

VPIRVGNTTSVQRVTTVNGVSNHRRPKRPFENLLCGINRGLVPRSTV"

join(5465..5512,5607..5654,5762..5786,6120..6194,

6385..6489,6572..6844)

/gene="At2g47640"

5465..6844

/gene="At2g47640"

/note="F17A22.3"

join(5650..5654,5762..5786,6120..6194,6385..6489,

6572..6688)

/gene="At2g47640"

/codon_start=1

/product="putative small nuclear ribonucleoprotein D2"

/protein_id="AAC63620.1"

/db_xref="GI:3738278"

/translation="MSKPMEDTNGKTEEBENTGTPLSVLMKSVKNNTQVLNCRNR

KLLGRVAFDRHCNMVLENVREMTVEVPTGKGGKALPVRNDRFISKMFRLGDSVII

VLRNPK"

complement(join(6904..7235,7321..7413,7965..8039,

8128..8211,8290..8369,8740..8887,8998..9714))

/gene="At2g47650"

complement(6904..9714)

/gene="At2g47650"

/note="F17A22.4"

complement(join(7008..7235,7321..7413,7965..8039,

8128..8211,8290..8369,8740..8887,8998..9621))

/gene="At2g47650"

/note="The first intron is U12-dependent (AT/AC class)"

/codon_start=1

/product="putative nucleotide-sugar dehydratase"

/protein_id="AAC63621.1"

/db_xref="GI:3738279"

/translation="MASELTNRHREIEQAESEYYPKPKPWFAIRPIRYMLRQRL

VEVLGATATLGTFTFSKSNHQPIDYDVPDLSYGMRSSESYLPATIIHKKPSLEYM

SRIGSAGGKIPGLKARKVLRVVVTGGAGFVGSGLVDRLMARGDNVIVVDNPTGRKEN

VNHNNFNEMIRHDVVEPILLEDVDQIYHLACPASPHYKFNPKTKITNVVGLTNM

LGAKRVGARELLTSTSEYGDLPQHPQVETWGNVNPFGVRSYDEGKRTAETLMD

YHRGANVEVRIARIENTYGPRLCIDDGRVSNVFAQALKEPLTVYDGKQTRSFQFY

SDLVGLMLRLEGGHGPENLGNPGFTMLELAKVQVETIDPNAKIEFRPNTDDPHK

RKPDITKAKELIGWEPKVALRQGLPLMKVDRQRVRFGQKQDSSTSSSTE"

7061..9447

/note="molecular marker 90J19T7"

/rpt_family="(GAA)n"

Join(<11209..11311,11413..>11417)

/gene="At2g47660"

<11209..>11417

/gene="At2g47660"

/note="F17A22.5; predicted by genscan"

Join(11209..11311,11413..11417)

CDS

misc_feature

repeat_region

mRNA

gene

RESULT	12
AC0114976	DNA
LOCUS	24766 bp HTG
DEFINITION	Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered pieces.
ACCESSION	AC014976 .
VERSION	AC014976.1 GI:6436359
KEYWORDS	HTG; HTGS_PHASE2.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 24766) Adams M and Venter J.C. Direct Submission Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	This sequence was identified as CDM:10213743 by the submitter.

AE003664 266340 bp DNA INV 04-OCT-2000
Drosophila melanogaster genomic scaffold 142000013386055 section 57
of 63, complete sequence.
AE003664 AE002690
AE003664.1 GI:7298621
HTG.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 266340)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J.J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Garq,N.S.,

Gelbart,W.M., Glasser,K., Glodek,A., Cong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
 Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarriy,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Swirskas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
 Weinstein,G.M., Wellenbach,J., Williams,S.M., Woodage,T.,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 266340)
 Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 Location/Qualifiers
 1. 266340
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2L"
 complement(join(542..567,703..1027,1121..1407,1597..2899,
 2958..3179,3470..3591,3725..>3906))
 /gene="CG10462"
 /product="CT29376"
 /db_xref="FLYBASE:FBan0010462"
 /db_xref="FLYBASE:FBgn0032815"
 complement(<542..>3906)
 /gene="CG10462"
 /map="38A2-38A2"
 /db_xref="FLYBASE:FBan0010462"
 /db_xref="FLYBASE:FBgn0032815"
 complement(join(542..567,703..1027,1121..1407,1597..2899,
 2958..3179,3470..3591,3725..3794))
 /gene="CG10462"
 /note="CG10462 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0010462"
 /db_xref="FLYBASE:FBgn0032815"
 /protein_id="AAF53838.1"
 /db_xref="GI:7298622"
 /translation="MEDIKMEPQDEVTSSIRLSACDTLKILDRGSMITVPIRD
 RLRSDEFLTASLEGHQIIFASEEDNPNKCDICQYKFSQDIISSRHVHNN
 LHFRLCGIWLDTLEIHEYRQHPGRLCCRCVDFWANSADOLAHMOKHIYV
 RFVCSQTRHFSKOSILNHRKQSDCILLDFKGSXSKPVLAVSTVPRESNCY
 VQKEEPLNADERMLEDNPNLVKLEQDEMPDYSIRKEQSGNSVWTAHTPL
 NKLPLALATTCOMHLKRFKFAINKSKSYIIITFSAGVRKGLLTINVESLNKLS
 NRSVDDANIPISNPFSSNIAANILKVLPSNCMVIKLPNPTKIFKTPGQAPAPK
 DSVPTVLTSGTNSNFNIKPTKITSISTVSGMGNPHASAKTVQPTSHRSSEFLVDSF
 NNSKSRPESKI INEFNRNIOIEKTLPLGTVLQSKVTMDSTSSPDAPKPELM
 LNPALMARELEIQPLYRMTWCPCLQRMFEKFAHRLTNKHLTDEKCNLVK
 VLMSKSTALSQVYNKGLSPISINIEKPKVSKPKNQSEQCTCLAVFTFGAL
 RHKSIHGTELPSCYCDKFRPTGGVRVHRRHRTGEKPKKICSLDTHRETLIS
 HLSRHGMKRYKCYGCDKFFVVSGLRAHRLRPDTCGVKFTFAHGRPRVIRGEV
 VFQVPHNGHVRLESDPLNLSQRDQNSTPETAIWMRGITPE"
 join(<4094..4166,4231..>4709)
 /gene="CG10447"
 /product="CT29328"
 /db_xref="FLYBASE:FBan0010447"
 /db_xref="FLYBASE:FBgn0032816"
 /note="CG10447 gene product"

gene

CDS

mRNA

gene

CDS

<4094..>4709
 /gene="CG10447"
 /map="38A2-38A2"
 /db_xref="FLYBASE:FBan0010447"
 /db_xref="FLYBASE:FBgn0032816"
 join(4094..4166,4231..4709)
 /gene="CG10447"
 /note="CG10447 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0010447"
 /db_xref="FLYBASE:FBgn0032816"
 /protein_id="AAF53839.1"
 /db_xref="GI:7298623"
 /translation="MSNEEDSQVYLVNMLVKEDEASDDSDKQGGIMLRQDRFLP
 ICNIIKMKVVPONGKIADKRECIQECVSEFISSEAIERSAENRKTVNGDDL
 LVAFNLGFDNVEPLSYLYQKYREVSSEITICTAVNNQVTHCSRTRNIVISFWTPVIR
 TTMGPPQPMKGNNSNPFUTYPS"
 complement(join(11183..11607,11666..17621,17682..20176,
 20236..21318,21392..21650,21710..22039,22098..22243,
 23832..>24830))
 /gene="CG10631"
 /product="CT29778"
 /db_xref="FLYBASE:FBan0010631"
 /db_xref="FLYBASE:FBgn0032817"
 complement(<11183..>24830)
 /gene="CG10631"
 /map="38A2-38A3"
 /db_xref="FLYBASE:FBan0010631"
 /db_xref="FLYBASE:FBgn0032817"
 complement(join(11183..11607,11666..17621,17682..20176,
 20236..21318,21392..21650,21710..22039,22098..22243,
 23832..24483))
 /gene="CG10631"
 /note="CG10631 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0010631"
 /db_xref="FLYBASE:FBgn0032817"
 /protein_id="AAF53840.1"
 /db_xref="GI:7298624"
 /translation="MSQQHPHHAHPHVAHYPPVTPMSLQQQQQQQQHQAQLSPQ
 QQQQHANNYSYVSYPTPHSPAPAPAPSKAANSGNNNNMGGVGGPGGQ
 AQGYGAGGLNVGAVGGGPNYGLGANTVAYAHNOLLOYQQQQQQQQQQQQQ
 QQHQHQLPQHSIQRPYGHNMHTGSTPIIKSEPMWATQQPNMAPPAPPEVLKS
 EPIDHSYKSNIDIDNTPADFSKSEFSEDSPLKVELTKDESGTSTFLRKLQ
 QDRGNESLQICQRCQKVEFKQVLRHVAESNCIQEYDFKSCIPKSEMTTELQR
 HKLHRADEFCHYCKGHEFTIAECEAHYMQHEYDFVCNMCSTGTFATRELQYHL
 POHQOQRDCPICRLWYOTALELHEHRLAAPYFCGYVTGGSSASQSOAQOHTN
 YLQDCHMATMPTPHKTTTPSGSSLPATAALNSLLQCOQANADGAAMFAAMKN
 EVNKMERSYNSSTSESSYQVDSGNNAGSDSSSMHAGIAGPAHSTLDDSDAL
 CCVPLCGVRKSTPTLOFFTFPKDEKYLNLHNLKMFHIIPAASTANFRICSMHFPKR
 CYNRYSLCYWAYPTNLGDDVANYLNRELNTFTTGEVARSMPHCTSQSGENLK
 FYNPKDKSLIKWCONARLPQAKPRHFCSEHFEERCIGFKRLKPWAVPTLHGAQ
 YGKIHDPKNLYVEEKRCNLCNFRSSSDNNMSLYRPPRDEVLLRWCYNRLDPGV
 YGKNHKICSAHFIEKALGLRLSPCAVPTLHGHDTFNIVYENLWPPPTAPNSHSS
 GLQHQQQSSQSDVCCVSCSKRNHNENITFTIPRQPMRKNKCNLKIPEKHKG
 GSGAMNASQMSQVSCSEKRNHNENITFTIPRQPMRKNKCNLKIPEKHKG
 MRICSLHFPIYICGMCRFVPTNLGDDDDIHRNDVIKRLNIRETCVAVCKRN
 RDRHANLHFRFNSVLLTKWGNLQRPVDPKICVHFEERLCRNLKLEK
 AVPTLSLGHENTIPYPLPTPEQVTEFSRPTAPNNGECCVETCKRNPVDDIKLY
 RPEEASVLAKWHLNQTSSOLTSMRNLHFEAHCIGKRMWAIPTNLNAGTLEN
 LYENPDSMLYKRRTHMKAKQASVYPTWVPCCLPHCRKVRALUNVOLYRPLNRS
 TLAKWHLNQLQVPMWGAQRRLCSAHFPHVLSKCPVPLATLNLNAPGLKIYQNP
 AKLKASKLQRCVIVESCRTAQQGVQLFRPHSTQVLQKWHNIIKTRPRAMRAQY
 RVSRRFRTEDDDMLKWNKNNKMPVDCQTVRNCNHFADCEHCVVEGEASKEQ
 PDLVPRFFTHSENGRLSAGATPLELGHDDDIYFNEAQAFADCEHCVVEGEASKEQ
 ELGHDDADLELVLPKPEDRYVDPVKCIVPTCQKTRRFRFDEQMSFKDANLFRWK
 HNLRLHSLFHRDERYKICNSHFEEDICIGKTLNIGSIPLELGHDDDLQVAPAE
 LOSNLFGRQRVHDESGGLIIEQFSESESDVKTDSDAKDENTRQVKLKTMSDLKCC
 VTRCSGRSLRLEHGARELPFTPGKQHLKWRHLNLEPDEVDSTRICSAHFNRRCDGK
 DLLEKWRANLRLAPDQCRICADHFEQVRGKLKLTGAVPTNLNIGHDEGLHYFPLE
 QLRSWAMPTQQLGHQEQPIYENPKIPGFTPTCALSHCRKRRSDINLRTYRPRSE
 IKVYMTREEGSSMTRLPKRELIDEEBEELAEPEHEHDMYDEKDHGFYFPLE
 LVETPAHPSDDAEYRGDEEDRDEDEDLDEAEHFPHPNDPPTPTPIRLREKPNANNV

repeat_region	/gene="At2g13480"	
	/note="T10F5.2; similar to GB:AAA75253"	
gene	/pseudo	
	complement(13418..13471)	
repeat_region	/rpt_family="(CAAAA)n"	
	14215..18083	
mRNA	/gene="At2g13490"	
	/note="T10F5.3"	
gene	/pseudo	
	complement(18097..18209)	
repeat_region	/rpt_family="(CAAAA)n"	
	join(<18363..19150,19298..>19536)	
mRNA	/gene="At2g13500"	
	<18363..>19526	
gene	/note="T10F5.4; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"	
	join(18363..19150,19298..19526)	
CDS	/gene="At2g13500"	
	/codon_start=1	
repeat_region	/product="putative Tall-like non-LTR retroelement protein"	
	/protein_id="AAD22675.1"	
mRNA	/db_xref="GI:4558657"	
	/translation="MVDLVDEIQNLGODYPLTFIPNEAYVMVEDNRNLSIARPL LRVONFNAVINALPRAGLTAHVHGRILDATYVOLLFOETIYVLSVQREVWLFNNM EVASORWEAPSNFVITIDLVQMRGIPILYVFEKTELEIAEGELGELISLDHFDATS TQIAYIRVRVIGITRLRFQRTFESGETALIRFQERLRIYSCFRFTHNLNYP YRQFQSDIRERAVFRDVLRSFNSQMTSSLOVPLHPPRPVPSGDATSGIMR YFEVGSRRGVQDGSNNMNLGESSKRNKLVILQDKESRQHQOETEQRMGGLPKR"	
gene	complement(join(<20428..20746,20999..>21534))	
	/gene="At2g13510"	
CDS	complement(<20428..>21534)	
	/gene="At2g13510"	
repeat_region	/note="T10F5.5; similar to GB:AAA75253; some members of this protein family have a weak CCHC zinc fingers that is mostly from retroviral gag proteins (nucleocapsid)"	
	complement(join(20428..20746,20999..21534))	
mRNA	/gene="At2g13510"	
	/codon_start=1	
gene	/product="putative Tall-like non-LTR retroelement protein"	
	/protein_id="AAD22676.1"	
CDS	/db_xref="GI:4558658"	
	/translation="MADKLDEIQNLGEPDPTLFTPEAYVMVAETNRLSIIARPL NRVONLSVAPALPNSWGLTFHVHGRVLDATYVQFLFHEMDLVSQREVWLFNNR EVASORWEAPSNFVITIDLVQMRGIPILYVFEKTELEIAEGELGELISLDHFDATS TQIAYIRVRVIGITRLRFQRTFESGETALIRFQERLRIYSCFRFTHNLNYP YRQFQSDIRERAVFRDVLRSFNSQMTSSLOVPLHPPRPVPSGDATSGIMR YFEVGSRRGVQDGSNNMNLGESSKRNKLVILQDKESRQHQOETEQRMGGLPKR"	
repeat_region	21534..21667	
	/rpt_family="(TAAAA)n"	
gene	complement(21749..25813)	
	/gene="At2g13520"	
CDS	/note="T10F5.6"	
	/pseudo	
repeat_region	complement(25902..26957)	
	/gene="At2g13530"	
mRNA	/note="T10F5.7"	
	/pseudo	
gene	join(<27480..27563,27648..27707,27807..27863,28349..28420,28529..28591,28765..28941,29126..29254,29478..29544,29724..29923,30139..30209,30672..30756,30903..30971,31063..31181,31489..31627,31829..32329,32420..32778,33154..33372)	
	/gene="At2g13540"	
CDS	<27480..33372	
	/gene="At2g13540"	
repeat_region	/note="T10F5.8"	
	join(27480..27563,27648..27707,27807..27863,28349..28420,28529..28591,28765..28941,29126..29254,29478..29544,29724..29923,30139..30209,30672..30756,30903..30971,31063..31181,31489..31627,31829..32329,32420..32778,33154..33372)	
	/gene="At2g13540"	
mRNA	/note="T10F5.9"	
	/pseudo	
gene	complement(34637..34727)	
	/rpt_family="(TA)n"	
CDS	complement(38932..40224)	
	/gene="At2g13550"	
repeat_region	complement(38932..40224)	
	/note="T10F5.9"	
mRNA	/gene="At2g13550"	
	/note="unknown protein"	
CDS	/codon_start=1	
	/protein_id="AAD22678.1"	
repeat_region	/db_xref="GI:4558660"	
	/translation="MTSIOIOTPLARMHLPTQOPNTRTGROKPSNHHHPDESSPS POEERHNPSEPTIRGIHQKTTSSSTANTQNTGCDKROPHIVNTDIASTDKTRS RSGNSKPOPHIQPKLKQSLKTCGAANNIETNLNRYYREIGPRTQHHQPYNHIKT RKKSSTDEIKRRTPEVLR"	
mRNA	join(<40516..40686,40766..40862,40943..41061,41171..41338,41499..41627,41718..41778,42210..42313,42390..42473,	
	41499..41627,41718..41778,42210..42313,42390..42473,	
CDS	Query Match 9.5%; Score 111.4; DB 12; Length 82403;	
	Best Local Similarity 62.3%; Pred. No. le-07;	
repeat_region	Matches 175; Conservative 0; Mismatches 106; Indels 0; Gaps 0;	
	QY 177 gacgagacggctgatcgccatcgcaacgctgatccgcatcatcgcgcgctgctgcg 236	
mRNA	Db 46811 GAACAAGACCGTTCTTCCCATTCGCAATGCGAAGGATCATGAAAAAGTTCTTCC 46752	
	QY 237 ggcaccccaagatctcgcgacgacgacgacgacgacgacgacgacgacgacgac 296	
CDS	Db 46751 GGTAACGGTAAGATCTCAAAAGACGCTTAAAGAACCGTTCAAGAAATGTGCTCGGAGTTC 46692.	
	QY 297 atcagcttcacgaggggagggagggagggagggagggagggagggagggagggaggg 356	
mRNA	Db 46691 ATTAGTTTCGTCTACTGTGAGCTTCTGACAGAGTCTCAAGAGAGAGAGAGAGAG 46632	
	QY 357 acgcccagagacgctgctgtggccatgagcgccctcgctcgacgactacgctcgagcgc 416	
CDS	Db 46631 AATGGAGATGATATCATTTGGGCTATCACAACCTCTCGGTTTCGAAGACTACGTTGGTCCA 46572	
	QY 417 ctcggccctacccctccaccgctaccgagctcgagggcga 457	
mRNA	Db 46571 TTAAGAGTCTACCTCTCAAAATATAGAGACCGAGGAGA 46531	
	RESULT 15	
repeat_region	ATHAP3A 832 bp mRNA PLN 12-SEP-1997	
	LOCUS Arabidopsis thaliana mRNA for Hap3a transcription factor.	
mRNA	DEFINITION Y13723	
	ACCESSION Y13723	
CDS	VERSION Y13723.1 GI:2398526	
	KEYWORDS hap3a gene; transcription factor.	
mRNA	SOURCE thale cress.	
	ORGANISM Arabidopsis thaliana	

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 832)

REFERENCE
AUTHORS
I (bases)
Edwards, D.

Direct Submission

Submitted (12-JUN-1997) Edwards D., Plant Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EA UK

REFERENCE

EDWARDS, D., SMITH, A.G. and MURRAY, J.A.
(cases 1 to 632)

Isolation and characterisation of CCAAT box binding proteins from higher plants

JOURNAL
Unpublished
manuscript plan

FEATURES

source

Location/Qualifiers

1.832

1. 832 /organism="Arabidopsis thaliana"

/organism="Arabidopsis"
/db xref="taxon:3702"

103. 55 /ab_xre

```
103. .528
/ gene="han3a"
```

```
103 /gene="r
```

```
103. 528
/geno="hen3"
```

```
/gene="P
```

```
/codon_start=1
/product="mycoproduct;on feat="
```

```
/product="Transcription"
```

```

/protein_id="CAA74051
/db_ref="GI:3309537"

```

```

/db_xref="GI:2398527"
/db_xref="CP009571.1"
/db_xref="CP009571.1"

```

```

/db_xref="SPTREMBL:O23633"
/translation="MADTPSSPACDGGESGVREQDRLPIANISRIMKKALPPNGK
IGKDAKDTVQECVSEFISETSTACDCKCQVPRDKVNCDDIISAMATLCTEDVITD

```

132 c 204 a 244 +

BASE CO
ORIGIN

Query Match 9.1%; Score 107; DB 13; Length 832;
Best Local Similarity 61.1%; Pred.No. 3.8e-06;
Matches 173; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

175 **acgagcaggaccgggctgataccgatacccaaacatcatccgcatcatccgagcccatcatccg**

173 GCYAGCAGGACCGGCTGATGCCGATCCGCACTGATCCGCATCATCGCGCGTGCTGC 234

db

161 GGGAGCAGGATCGATACCTTCCTATAGCTAATACTACCCGCAATCATCAACAACCTTCCC

161 GGGAGCAGGAATCGATACCTTCCCTATAGCTAATATCAGCAGGATCATGAAGAAAGCGTTGC 220

235 ccgcgacgccaaatctccagcggaccgaaggcgccggttccttccg

235 cggcgacgccaagatctcggacgacgacgccaaggagacgatccaggagtgcgtgtcggagt 294

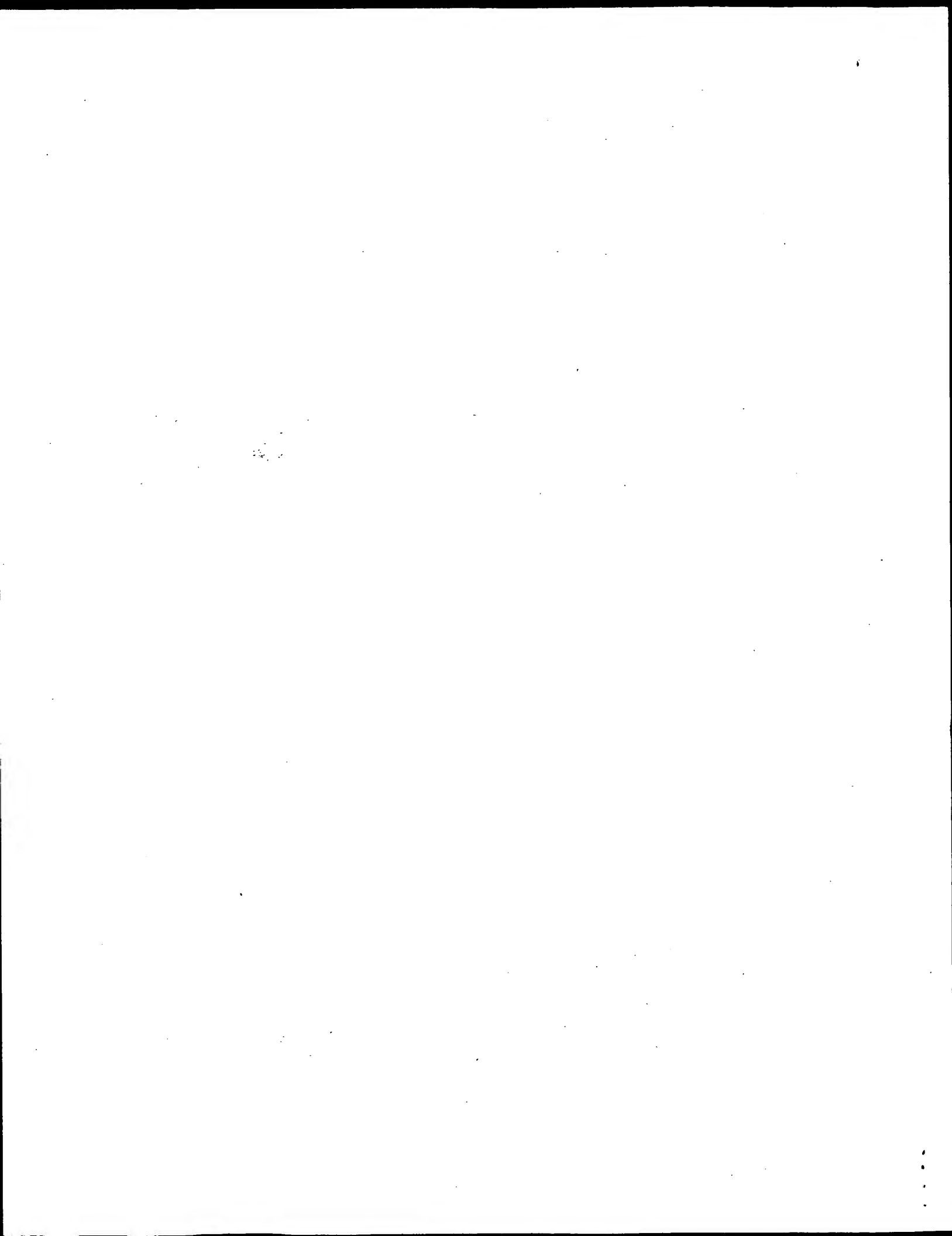
[illegible]

221 CTCCTAATGGTAAGATTGGAAAAGATGCTAAGGATACAGTTCAGGAATGCGTCTCTGAGT 280

[illegible]

295 acatcagcttcacgcgggggagggccaacgagcgggtgccagcgggagcagcgaagacca 354

Search completed: October 25, 2001, 12:05:47
Job time: 4682 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:19:30 ; Search time 12.26 Seconds
(without alignments)
466.894 Million cell updates/sec

Title: US-09-435-054-2

Perfect score: 1514

Sequence: 1 MDSSFLPAAGANGSAGG.....ASHTPQSGGLEHHPHAYK 278

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PTCUS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	494	32.6	208	4	US-09-103-478-2
2	403	26.6	90	4	US-09-103-478-19
3	341	22.5	90	4	US-09-103-478-20
4	305	20.1	90	4	US-09-103-478-22
5	301	19.9	90	4	US-09-103-478-21
6	301	19.9	90	4	US-09-103-478-24
7	301	19.9	90	4	US-09-103-478-25
8	298	19.7	90	4	US-09-103-478-26
9	295.5	19.5	144	4	US-09-319-989-4
10	270	17.8	90	4	US-09-103-478-28
11	269	17.8	85	4	US-09-103-478-27
12	262	17.3	90	4	US-09-103-478-29
13	208	13.7	57	4	US-09-103-478-23
14	160	10.6	156	1	US-08-681-812-7
15	147	9.7	1185	4	US-08-041-886-23
16	140.5	9.3	420	2	US-08-845-998-8
17	140.5	9.3	420	4	US-09-206-537-8
18	129.5	8.6	918	4	US-09-041-886-11
19	120	7.9	1958	1	US-07-945-283-2
20	118	7.8	738	3	US-08-864-038A-3
21	118	7.8	907	2	US-09-010-928B-4
22	114	7.5	410	4	US-09-431-573-5
23	112	7.4	553	3	US-09-083-351-2
24	112	7.4	553	4	US-09-083-352-2
25	112	7.4	745	2	US-09-010-928B-28
26	112	7.4	870	2	US-09-010-928B-2
27	109	7.2	410	4	US-09-431-573-4

28	107	7.1	410	1	US-08-123-343A-5	Sequence 5, Appli
29	107	7.1	410	1	US-08-123-343A-7	Sequence 7, Appli
30	106	7.0	340	5	PCT-US96-02331-13	Sequence 13, Appl
31	106	7.0	842	5	PCT-US96-02331-15	Sequence 15, Appl
32	105.5	7.0	2100	1	US-08-477-509B-80	Sequence 80, Appl
33	105.5	7.0	2100	3	US-08-482-085B-80	Sequence 80, Appl
34	105.5	7.0	2107	1	US-08-175-155-45	Sequence 45, Appl
35	105.5	7.0	2107	2	US-08-707-237A-51	Sequence 51, Appl
36	104.5	6.9	330	2	US-08-712-948-2	Sequence 2, Appli
37	104	6.9	333	2	US-08-712-948-1	Sequence 1, Appli
38	103.5	6.8	1057	3	US-08-931-820-1	Sequence 1, Appli
39	103.5	6.8	1341	3	US-08-963-825-18	Sequence 18, Appl
40	103	6.8	671	3	US-09-121-321-16	Sequence 16, Appl
41	103	6.8	671	4	US-08-933-803A-16	Sequence 16, Appl
42	103	6.8	1038	1	US-07-609-716-36	Sequence 36, Appl
43	103	6.8	1038	4	US-08-475-411A-36	Sequence 36, Appl
44	103	6.8	1038	4	US-08-478-029A-36	Sequence 36, Appl
45	102.5	6.8	595	1	US-08-425-069-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-103-478-2
; Sequence 2, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 208 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-103-478-2

Query Match 32.6%; Score 494; DB 4; Length 208;
Best Local Similarity 47.0%; Pred. No. 2.3e-38;
Matches 117; Conservative 19; Mismatches 47; Indels 66; Gaps 9;
QY 16 SAAGGANNAGAAQHAAPAIREDRLMPTIANVIRIMRVLPAHAKISDDAKETIQECVSE 75
DB 8 AGAGDKNGIVVQQPPCAREQDQYMPIANVIRIMRKTLPSPHAKISDDAKETIQECVSE 67
QY 76 YISFTIGNEANCOREORTITAEADVWMSRLGFDYVEPLGAYLHRYREFEGDARGVG 135
DB 68 YISFTIGNEANCOREORTITAEADVWMSRLGFDYVEPLGAYLHRYREFEGDARGVG 126
QY 136 LVPGAAPSRGGDHPSMSPAAMLSKRGVPSGAAMLPHHHHHHDMQMHAAHYGGTAVPPP 195
DB 127 L-----RG-----EPPSLRQT-----YGGNGI--- 143
QY 196 AGPPHGGFLPHPGSSHYL-----PYAYEPTYGGEHMAAAYGGAAYAPNGGSGDGS 251
DB 144 -----GF-----HGPSHGLPPPGPYG-----GMLDQSMVMGSGGGRYYQ-NGSSGQDES 185
QY 252 SCGGGGSAS 260
DB 186 SVGGGSSS 194

RESULT 2
US-09-103-478-19
; Sequence 19, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide

US-09-103-478-19
Query Match 26.6%; Score 403; DB 4; Length 90;
Best Local Similarity 82.2%; Pred. No. 2e-30;
Matches 74; Conservative 11; Mismatches 5; Indels 0; Gaps 0;
QY 36 REQDRLMPIANVIRIMRVLPAHAKISDDAKETIQECVSEYISFTIGNEANCOREQSKRT 95
DB 1 REQDYMPIANVIRIMRKTLPSPHAKISDDAKETIQECVSEYISFTIGNEANCOREQSKRT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
DB 61 ITAEDILWMSKLGDFQYVDPDLTVFINRYR 90
RESULT 3
US-09-103-478-20
; Sequence 20, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 90 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-103-478-20
Query Match 22.5%; Score 341; DB 4; Length 90;
Best Local Similarity 65.6%; Pred. No. 1.1e-24;
Matches 59; Conservative 19; Mismatches 12; Indels 0; Gaps 0;
QY 36 REQDRLMPIANVIRIMRVLPAHAKISDDAKETIQECVSEYISFTIGNEANCOREQSKRT 95

Db 1 REQDRLPIANIRIMKATPANGKIAKAKETVQECVSEFISFITSEASDKQREKRT 60
 QY 96 ITAEDVLWAMSRGLGFDYVEPLGAYLHRYR 125
 Db 61 INGDLLWAMATLGFEDYIEPLKYLQKYR 90

RESULT 4
 US-09-103-478-22
 ; Sequence 22, Application US/09103478
 ; Patent No. 6235975
 ; GENERAL INFORMATION:
 ; APPLICANT: Harada, John
 ; APPLICANT: Lotan, Tamar
 ; APPLICANT: Ohto, Masa-aki
 ; APPLICANT: Goldberg, Robert B.
 ; APPLICANT: Fischer, Robert L.
 ; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/103,478
 ; FILING DATE: 24-JUN-1998
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/026,221
 ; FILING DATE: 19-FEB-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/804,534
 ; FILING DATE: 21-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einhorn, Gregory P.
 ; REGISTRATION NUMBER: 38,440
 ; REFERENCE/DOCKET NUMBER: 023070-077611US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 22:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 90 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-103-478-22

Query Match 20.1%; Score 305; DB 4; Length 90;
 Best Local Similarity 64.4%; Pred. No. 2.3e-21;
 Matches 58; Conservative 13; Mismatches 19; Indels 0; Gaps 0;

QY 36 REQDRLPIANVIRIMRRVLPAAKISDDAKETIQECVSEYISFITGEANERCOREKRT 95
 Db 1 REQDIYLPANVARIMKTAIPQGTAKDAKCEVCVSEFISFITSEASERCHQEKRT 60
 QY 96 ITAEDVLWAMSRGLGFDYVEPLGAYLHRYR 125
 Db 61 INGDILFAMSTLGFQSYVEPLKYLQKYR 90

RESULT 5
 US-09-103-478-21

; Sequence 21, Application US/09103478
 ; Patent No. 6235975
 ; GENERAL INFORMATION:
 ; APPLICANT: Harada, John
 ; APPLICANT: Lotan, Tamar
 ; APPLICANT: Ohto, Masa-aki
 ; APPLICANT: Goldberg, Robert B.
 ; APPLICANT: Fischer, Robert L.
 ; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Townsend and Townsend and Crew LLP
 ; STREET: Two Embarcadero Center, Eighth Floor
 ; CITY: San Francisco
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/103,478
 ; FILING DATE: 24-JUN-1998
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/026,221
 ; FILING DATE: 19-FEB-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/804,534
 ; FILING DATE: 21-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einhorn, Gregory P.
 ; REGISTRATION NUMBER: 38,440
 ; REFERENCE/DOCKET NUMBER: 023070-077611US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 21:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 90 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-09-103-478-21

Query Match 19.9%; Score 301; DB 4; Length 90;
 Best Local Similarity 63.3%; Pred. No. 5.5e-21;
 Matches 57; Conservative 13; Mismatches 20; Indels 0; Gaps 0;

QY 36 REQDRLPIANVIRIMRRVLPAAKISDDAKETIQECVSEYISFITGEANERCOREKRT 95
 Db 1 REQDIYLPANVARIMKTAIPQGTAKDAKCEVCVSEFISFITSEASERCHQEKRT 60
 QY 96 ITAEDVLWAMSRGLGFDYVEPLGAYLHRYR 125
 Db 61 INGDILFAMSTLGFQSYVEPLKYLQKYR 90

RESULT 6
 US-09-103-478-24
 ; Sequence 24, Application US/09103478
 ; Patent No. 6235975
 ; GENERAL INFORMATION:
 ; APPLICANT: Harada, John
 ; APPLICANT: Lotan, Tamar
 ; APPLICANT: Ohto, Masa-aki
 ; APPLICANT: Goldberg, Robert B.
 ; APPLICANT: Fischer, Robert L.
 ; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses

```

7
RESULT
US-09-103-478-25
; Sequence 25, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

```

36	REQDRKPLMIANVIMRVLPAHAKISDDAKETIQBCVSEYISFTTGEANERCOEQRKT	95
QY		
db		
1	REQDIYLPDIANVARIMKNAITPGTKIAKDACEVCQBCSEFISFTTSEASERCHOEKRRKT	60
QY		
96	ITAEVDLWAMSRLEGFDYVEPGLAYLHRYR	125
QY		
61	INGRDIPLFAMSTLGFQSVYRPLKLYLQKFR	90
db		

```

RESULT      8
US-09-103-478-26
; Sequence 26, Application US/09103478
; Patent No. 6235975
;
; GENERAL INFORMATION:
;
; APPLICANT: Harada, John
;
; APPLICANT: Lotan, Tamar
;
; APPLICANT: Ohto, Masa-aki
;
; APPLICANT: Goldberg, Robert B.
;
; APPLICANT: Fischer, Robert L.
;
; TITLE OF INVENTION:  LIBRARY COTYLEDON1 Genes and Their Uses
;
; NUMBER OF SEQUENCES:  29
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE:  Townsend and Townsend and Crew LLP
;
; STREET:  Two Embarcadero Center, Eighth Floor
;
; CITY:  San Francisco
;
; STATE:  California
;
; COUNTRY:  USA
;
; ZIP:  94111-3834
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE:  Floppy disk
;
; COMPUTER:  IBM PC compatible
;
; OPERATING SYSTEM:  PC-DOS/MS-DOS
;
; SOFTWARE:  PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER:  US/09/103,478
;
; FILING DATE:  24-JUN-1998
;
; CLASSIFICATION:  800
;
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-26

Query Match 19.7%; Score 298; DB 4; Length 90;
Best Local Similarity 61.1%; Pred. No. 1e-20;
Matches 55; Conservative 19; Mismatches 16; Indels 0; Gaps 0;
QY 36 REQDRLPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQREQRKT 95
Db 1 KEQDWLPPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQREQRKT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
Db 61 VNGEDILFAMTSLGFENYAEVLYLAKYR 90

RESULT 9
US-09-319-989-4
Sequence 4, Application US/09319989
Patent No. 6190914
GENERAL INFORMATION:
APPLICANT: Grivell, Leslie A.
APPLICANT: Teixeira de Mattos, Maarten J.
TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
TITLE OF INVENTION: METHODS
FILE REFERENCE: 24615-20123.00
CURRENT APPLICATION NUMBER: US/09/319,989
EARLIER FILING DATE: 1999-06-14
EARLIER FILING DATE: 1997-12-12
EARLIER FILING DATE: 1997-12-12
EARLIER FILING DATE: 1996-12-12
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 144
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: HAP3
US-09-319-989-4

Query Match 19.5%; Score 295.5; DB 4; Length 144;
Best Local Similarity 50.4%; Pred. No. 3.2e-20;
Matches 60; Conservative 21; Mismatches 33; Indels 5; Gaps 2;
QY 8 PAAGENGAAGGANGGAQQAAPAREQDRLMPIANVIRMRVLPAAKISDDAKE 67
Db 13 PEDTQENG--GNASSSGSLQQ--ISTREQDWLPINNVARLMKNTLPPSAKYSKDAKE 67
QY 68 TIQECVSEYISFITGEANERCQREQRKTITAEDVLWMSRLGFDYVEPLGAYLHRYRE 126

Db 68 CMQECVSELISFVTSEASDRCADKRRKTINGEDILISLHALGFENYAEVLYLAKYRQ 126
RESULT 10
US-09-103-478-28
Sequence 28, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-28

Query Match 17.8%; Score 270; DB 4; Length 90;
Best Local Similarity 56.7%; Pred. No. 4e-18;
Matches 51; Conservative 16; Mismatches 23; Indels 0; Gaps 0;

QY 36 REQDRLMPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQREQRKT 95
Db 1 REQDWLPINNVARLMKNTLPPSAKYSKDAKECMQECVSELISFVTSEASDRCADKRRKT 60
QY 96 ITAEDVLWMSRLGFDYVEPLGAYLHRYR 125
Db 61 INGEDILISLHALGFENYAEVLYLAKYR 90

RESULT 11
US-09-103-478-27
Sequence 27, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:

APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-29

Query Match 17.8%; Score 262; DB 4; Length 90;
Best Local Similarity 56.2%; Pred. No. 2.2e-17;
Matches 50; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

QY 37 EODRLPIANVIRMRVLPAAHAKISDDAKETIQCVSEYISFITGEANERCQREKRTI 96
Db 2 EQDRWLPINNVARLMKNTLPATTKVSKDAKCMQCVSEFISFVTSEACDRCTSGKRTI 61
QY 97 TAEDVLWAMSRIGFDDYVEPLGAYLHYR 125
Db 62 NGEDILLSLHALGFENYAEVLKYLAKYR 90

RESULT 13
US-09-103-478-23
Sequence 23, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 85 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-478-27

Query Match 17.8%; Score 269; DB 4; Length 85;
Best Local Similarity 60.0%; Pred. No. 4.6e-18;
Matches 51; Conservative 18; Mismatches 16; Indels 0; Gaps 0;

QY 41 LPIANVIRMRVLPAAHAKISDDAKETIQCVSEYISFITGEANERCQREKRTI 100
Db 1 LPIANVIRMRVLPAAHAKISDDAKETIQCVSEYISFITGEANERCQREKRTI 60
QY 101 VLWAMSRIGFDDYVEPLGAYLHYR 125
Db 61 VLLALNTLGFENYAEVLKISLTKYR 85

RESULT 12
US-09-103-478-29
Sequence 29, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP

; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/103,478
 ; FILING DATE: 24-JUN-1998
 ; CLASSIFICATION: 800
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 09/026,221
 ; FILING DATE: 19-FEB-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/804,534
 ; FILING DATE: 21-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einhorn, Gregory P.
 ; REGISTRATION NUMBER: 38,440
 ; REFERENCE/DOCKET NUMBER: 023070-077611US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (415) 576-0200
 ; TELEFAX: (415) 576-0300
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 57 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-09-103-478-23

Query Match 13.7%; Score 208; DB 4; Length 57;
 Best Local Similarity 66.7%; Pred. No. 1.2e-12;
 Matches 38; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
 QY 69 IQECVSEVIFITGEANERCORQRTITAEVDLWMSRIGFDDYVEPLGAYLHRYR 125
 Db 1 VQECVSEVIFITGEANERCORQRTITAEVDLWMSRIGFDDYVEPLGAYLHRYR 57

RESULT 14
 US-08-681-812-7
 ; Sequence 7, Application US/08681812
 ; Patent No. 5763593
 ; GENERAL INFORMATION:
 ; APPLICANT: Young, Richard A.
 ; APPLICANT: Gadbois, Ellen L.
 ; APPLICANT: Chao, David M.
 ; TITLE OF INVENTION: TBP-Associated Global Negative Regulator
 ; TITLE OF INVENTION: and Methods of Use Thereof
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
 ; STREET: Two Militia Drive
 ; CITY: Lexington
 ; STATE: Massachusetts
 ; COUNTRY: USA
 ; ZIP: 02173
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/681,812
 ; FILING DATE: 29-JUL-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Granahan, Patricia
 ; REGISTRATION NUMBER: 32,227
 ; REFERENCE/DOCKET NUMBER: WHI96-07
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 617-861-6240
 ; TELEFAX: 617-861-9540
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 156 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-681-812-7

Query Match 10.6%; Score 160; DB 1; Length 156;
 Best Local Similarity 31.5%; Pred. No. 1.2e-07;
 Matches 34; Conservative 28; Mismatches 38; Indels 8; Gaps 2;
 QY 37 EQDLMPIANVIRIMRVLPAHAKISDDAKETIOECVSEVIFITGEANERCORQRTI 96
 Db 4 DDDLTPRAAINKMIKETLP-NVRVANDARELVVNCCTEFILHLSSEANEICNKSEKRTI 62
 QY 97 TAEVDLWMSRIGFDDYVEPLGAYLHRYREFEGDARGVGLVPCAAPSR 144
 Db 63 SPEHVIALESIGF-----GSYISEVKEVLQECKTVALKRRKASSR 103

RESULT 15
 US-09-041-886-23
 ; Sequence 23, Application US/09041886
 ; Patent No. 6235872
 ; GENERAL INFORMATION:
 ; APPLICANT: Bredesen, Dale E.
 ; APPLICANT: Rabizadeh, Sharoz
 ; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
 ; TITLE OF INVENTION: Polypeptides and Methods of Use
 ; NUMBER OF SEQUENCES: 72
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell & Flores LLP
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: United States
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/041,886
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-LJ 2626
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1185 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-09-041-886-23

Query Match 9.7%; Score 147; DB 4; Length 1185;
 Best Local Similarity 31.7%; Pred. No. 2.7e-05;
 Matches 45; Conservative 11; Mismatches 56; Indels 30; Gaps 7;
 QY 149 HPHMSPA--AMLKSRGVPVGAAMLPHHHHHHDMQMAAM-----YGTAVPPPPGPPH 200
 Db 458 HPGFPFPGSTGAQSTAHPPVS-----THHHHHQQQQQQQQQQHHGNSGPPPPGPPH 512
 QY 201 HGGFLMPHPQSSHYL-PYAYEPTYGGHMAAYYGAAYAPNGNGSGDGSGGG--- 256

Db 513 -----PLEGGSSHHAPYAMSPSLG---SLRYPYPGPAHLPPPHSQVYSQAGNGPPV 563
Qy 257 -----GSASHTPOGSGGLEHHP 274
 |:| | || |
Db 564 SSSNSSSTSQGSYPCHSP 585

Search completed: October 25, 2001, 09:21:43
Job time: 133 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:16:40 ; Search time 20.67 Seconds
(without alignments)
815.359 Million cell updates/sec

Title: US-09-435-054-2

Perfect score: 1514

Sequence: 1 MDSSFLPAAGSAAG.....ASHTPQSGGLEHPPHPPAYK 278

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0601.*
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1514	100.0	278	21	Maize LEC1 # 1. Z
2	516.5	34.1	355	21	Soybean LEC1 # 2 p
3	516.5	34.1	373	21	Soybean LEC1 # 2 p
4	508	33.6	240	21	Soybean LEC1 # 1.
5	494	32.6	208	19	Arabidopsis leafy-
6	494	32.6	208	21	Amino acid sequenc
7	489.5	32.3	280	21	Wheat LEC1. Tril
8	484	32.0	214	21	Veronica mespilifol
9	444	29.3	146	21	Argemone mexicana
10	443	29.3	171	21	Soybean LEC1 # 3.
11	420.5	27.8	219	21	Pinus radiata tran

12	409	27.0	90	21	Arabidopsis LEC1 c
13	376.5	24.9	141	21	Arabidopsis thalia
14	360.5	23.8	138	21	Arabidopsis thalia
15	360.5	23.8	139	21	Arabidopsis thalia
16	324.5	21.4	207	12	NP-YB. Homo sapie
17	324.5	21.4	210	21	Arabidopsis thalia
18	309.5	20.4	206	21	Human prostate can
19	303	20.0	107	21	Maize LEC1 # 3. Z
20	303	20.0	107	21	Arabidopsis thalia
21	297	19.6	108	21	Arabidopsis thalia
22	293.5	19.4	125	21	Pinus radiata tran
23	287	19.0	104	21	Eucalyptus grandis
24	287	19.0	105	21	Pinus radiata tran
25	235	15.5	65	21	Pinus radiata tran
26	226	14.9	94	21	Arabidopsis thalia
27	200	13.2	88	21	Plant LEC1 consens
28	199	13.1	78	21	Eucalyptus grandis
29	197	13.0	160	21	Pinus radiata tran
30	197	13.0	163	21	Eucalyptus grandis
31	192	12.7	160	21	Arabidopsis thalia
32	185.5	12.3	155	21	Maize LEC1 # 2. Z
33	185.5	12.3	158	21	Arabidopsis thalia
34	182.5	12.1	400	20	Arabidopsis thalia
35	181	12.0	143	21	Human c-Maf protei
36	173	11.4	117	21	Arabidopsis thalia
37	171	11.3	93	21	Arabidopsis thalia
38	171	11.3	96	21	Pinus radiata tran
39	171	11.3	98	21	Zea mays protein f
40	171	11.3	98	20	Zea mays protein f
41	169.5	11.2	153	21	Soybean Drl protei
42	165.5	10.9	186	22	Zea mays protein f
43	153	10.1	113	20	Arabidopsis thalia
44	147	9.7	1185	20	Human breast cance
45	144	9.5	93	21	Wheat Drl protein
					Human atrophin I p
					Human secreted pro

ALIGNMENTS

RESULT 1

AA196214
ID AAY96214 standard; Protein: 278 AA.

XX AAY96214;

XX 11-SEP-2000 (first entry)

XX Maize LEC1 # 1.

XX Maize; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
KW selectable marker; transgenic plant; transgenic seed; HAP3.
XX Zea mays.

XX WO200028058-A2.

XX 18-MAY-2000.

XX 09-NOV-1999; 99WO-US26514.

XX 09-NOV-1998; 98US-0107643.

XX 10-NOV-1998; 98US-0107810.

XX (PION-) PIONEER HI-BRED INT INC.
XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
XX Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;

XX WPI; 2000-376568/32.

XX N-PSDB; AAA27450.

XX New HAP3-type CCAAT-box binding transcriptional activators,

PT particularly leafy cotyledon 1 transcriptional activator, useful for
 XX inducing somatic embryogenesis or apomixis in a plant cell -
 PS Claim 13; Page 77-78; 94pp; English.
 XX
 CC The present sequence is the maize leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs, i.e.
 CC the replacement of sexual reproduction by asexual reproduction, LEC1
 CC expression in the nucleellus integument, or cell specific expression in
 CC the megaspore mother cell would trigger embryo formation from maternal
 CC tissues only. This results in the production of seeds identical to the
 CC parent. Using LEC1, transgenic high yielding seeds could be developed. In
 CC addition, LEC1 could be used for positive selection of a transformed cell
 CC (transgenic plant), for increasing transformation efficiency and for
 CC increasing recovery of regenerated plants.
 XX
 SQ Sequence 278 AA;

Query Match 100.0%; Score 1514; DB 21; Length 278;
 Best Local Similarity 100.0%; Pred. No. 3.3e-127;
 Matches 278; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSSSFLPAAGAGSAGGANNNGAAQQAAPAIHQDRLMPIANVIRMRVLPAAHAK 60
 DB 1 MDSSSFLPAAGAGSAGGANNNGAAQQAAPAIHQDRLMPIANVIRMRVLPAAHAK 60
 QY 61 ISDDAKETIQECVSEYISFTITGANERCQRTTAEEDVLWMSRLGDFDYVEPLGAY 120
 DB 61 ISDDAKETIQECVSEYISFTITGANERCQRTTAEEDVLWMSRLGDFDYVEPLGAY 120
 QY 121 LHRYREFEGDARGVGLVGAAPSRGDDHPSMSPAAMLKSRGVSQAAMLPHHHHHHDM 180
 DB 121 LHRYREFEGDARGVGLVGAAPSRGDDHPSMSPAAMLKSRGVSQAAMLPHHHHHHDM 180
 QY 181 QMHAAMYGGTAVPPAGPHHGGFLMPHFGSSHYLPYAEPTYGGEHAMAAYYGAAYA 240
 DB 181 QMHAAMYGGTAVPPAGPHHGGFLMPHFGSSHYLPYAEPTYGGEHAMAAYYGAAYA 240
 QY 241 PNGSGDGGSGGGSGASHTPQSGGLEHPPHFAK 278
 DB 241 PNGSGDGGSGGGSGASHTPQSGGLEHPPHFAK 278

RESULT 2
 AAY96224
 ID AAY96224 standard; Protein; 355 AA.
 XX
 AC AAY96224;
 XX
 11-SEP-2000 (first entry)
 XX
 DE Soybean LEC1 # 2 protein # 2.
 XX
 KW Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX
 OS Glycine max.
 XX
 PN WO200028058-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-US26514.
 XX
 PR 09-NOV-1998; 98US-0107643.
 PR 10-NOV-1998; 98US-0107810.
 XX
 XX (PION-) PIONEER HI-BRED INT. INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E. I.
 XX

PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 XX Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 DR WPI: 2000-376568/32.
 DR N-PSDB; AAA27460.
 XX
 PT New HAP3-type CCAAT-box binding transcriptional activators,
 PT particularly leafy cotyledon 1 transcriptional activator, useful for
 PT inducing somatic embryogenesis or apomixis in a plant cell -
 XX
 XX Claim 13; Page 89-90; 94pp; English.
 XX
 CC The present sequence is the soybean leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs,
 CC i.e. the replacement of sexual reproduction by asexual reproduction,
 CC LEC1 expression in the nucleellus integument, or cell specific expression
 CC in the megaspore mother cell would trigger embryo formation from
 CC maternal tissues only. This results in the production of seeds identical
 CC to the parent. Using LEC1, transgenic high yielding seeds could be
 CC developed. In addition, LEC1 could be used for positive selection of a
 CC transformed cell (transgenic plant), for increasing transformation
 CC efficiency and for increasing recovery of regenerated plants.
 CC Note: this sequence is different from that of AAY96220, even though
 CC they are both encoded by the same nucleotide sequence (AAA27460). The
 CC AAY96220 sequence has 18 stop codons distributed throughout the CDS,
 CC while the protein of AAY96224 has the stop codons omitted.
 XX
 SQ Sequence 355 AA;

Query Match 34.1%; Score 516.5; DB 21; Length 355;
 Best Local Similarity 53.8%; Pred. No. 3.3e-38;
 Matches 113; Conservative 16; Mismatches 50; Indels 31; Gaps 4;

QY 16 SNAGGANNNGAAQQAAPAIHQDRLMPIANVIRMRVLPAAHAKISDDAKETIQECVSE 75
 DB 38 ssdqncshsaageenectvreqdfrfmpianvirimrkilpphakisdaketi qecvse 97
 QY 76 YISFTITGANERCQRTTAEEDVLWMSRLGDFDYVEPLGAYLHRYREFEGDARGVG 135
 DB 98 yisfitgeanercqrktitaedvlwamsklgfdyiepltmlylhryrelegdrtsm- 156
 QY 136 LVPGAAPSRGDDHPSMSPAAMLKSRGVSQAAMLPHHHHHHDMQMAHYGGTAV--- 192
 DB 157 -----rgeplgkrtveyatl-----atafppppfhhhngyfgaampmgtyvret 200
 QY 193 PPPAGPHHGGFLMPHFGSSHYLPYAEVPEP 222
 DB 201 ppnaasshh-----hhgisnahep 219

RESULT 3
 AAY96220
 ID AAY96220 standard; Protein; 373 AA.
 XX
 AC AAY96220;
 XX
 11-SEP-2000 (first entry)
 XX
 DE Soybean LEC1 # 2 protein # 1.
 XX
 KW Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX
 OS Glycine max.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 1...373
 FT /note= "All Xs are unspecified residues encoded by
 FT stop codons"
 XX

PN WO200028058-A2.
 XX 18-MAY-2000.
 PD 09-NOV-1999; 99WO-US26514.
 XX 09-NOV-1998; 98US-0107643.
 XX 10-NOV-1998; 98US-0107810.
 PR (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX WPI; 2000-376568/32.
 DR N-PSDB; AAA27460.
 XX New HAP3-type CCAAT-box binding transcriptional activators,
 PT particularly leafy cotyledon 1 transcriptional activator, useful for
 PT inducing somatic embryogenesis or apomixis in a plant cell -
 XX Claim 13; Page 89-90; 94pp; English.
 PS The present sequence is the soybean leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs,
 CC i.e. the replacement of sexual reproduction by asexual reproduction,
 CC LEC1 expression in the nucleellus integument, or cell specific expression
 CC in the megaspore mother cell would trigger embryo formation from
 CC maternal tissues only. This results in the production of seeds identical
 CC to the parent. Using LEC1, transgenic high yielding seeds could be
 CC developed. In addition, LEC1 could be used for positive selection of a
 CC transformed cell (transgenic plant), for increasing transformation
 CC efficiency and for increasing recovery of regenerated plants.
 CC Note: this sequence is different from that of AAY96224, even though
 CC they are both encoded by the same nucleotide sequence (AAA27460). The
 CC AAY96220 sequence has 18 stop codons distributed throughout the CDS,
 CC while the protein of AAY96224 has the stop codons omitted.
 XX Sequence 373 AA;
 SQ
 Query Match 34.1%; Score 516.5; DB 21; Length 373;
 Best Local Similarity 53.8%; Pred. NO. 3.5e-38;
 Matches 113; Conservative 16; Mismatches 50; Indels 31; Gaps 4;
 QY 16 SAAGGANGGAQAQAHPAIREQDRLMPTANVIRIMRRLVPAHAKISDDAKETIQECVSE 75
 Db 38 ssdncnshsaageenectvreqdrfmpianvirimrkilpphakisddaketi qcvse 97
 QY 76 YISFITGEANERCQQRKTTAEDVLWAMSRGLGFDYVEPLGAYLHRYREPEGDARGVG 135
 Db 98 yisfitgeanercqrkitaedvlwamskigfdyieplumylhryrelegdrtsm- 156
 QY 136 LVPGAAPSRGDHHPHSMSPAMKSRGPVSGAAMLPHHHHHHDMQMHAAWYGGTAV--- 192
 Db 157 -----rgeplgkrtyeatl-----atafvppfhhngyfgaampmgtyvret 200
 QY 193 PPPAGPPHGGFLMPPHPOGSSHYLPYAYEP 222
 Db 201 ppaasshh-----hhglsnahep 219
 RESULT 4
 AAY96216
 ID AAY96216 standard; Protein; 240 AA.
 XX AAY96216;
 XX
 XX 11-SEP-2000 (first entry)
 XX

DE Soybean LEC1 # 1.
 XX Soybean; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX Glycine max.
 OS WO200028058-A2.
 PN 18-MAY-2000.
 PD 09-NOV-1999; 99WO-US26514.
 XX 09-NOV-1998; 98US-0107643.
 PR 10-NOV-1998; 98US-0107810.
 XX (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX WPI; 2000-376568/32.
 DR N-PSDB; AAA27456.
 XX New HAP3-type CCAAT-box binding transcriptional activators,
 PT particularly leafy cotyledon 1 transcriptional activator, useful for
 PT inducing somatic embryogenesis or apomixis in a plant cell -
 XX Claim 13; Page 82; 94pp; English.
 PS The present sequence is the soybean leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformants. When apomixis occurs, i.e.
 CC the replacement of sexual reproduction by asexual reproduction, LEC1
 CC expression in the nucleellus integument, or cell specific expression in
 CC the megaspore mother cell would trigger embryo formation from maternal
 CC tissues only. This results in the production of seeds identical to the
 CC parent. Using LEC1, transgenic high yielding seeds could be developed. In
 CC addition, LEC1 could be used for positive selection of a transformed cell
 CC (transgenic plant), for increasing transformation efficiency and for
 CC increasing recovery of regenerated plants.
 XX Sequence 240 AA;
 SQ
 Query Match 33.6%; Score 508; DB 21; Length 240;
 Best Local Similarity 52.5%; Pred. NO. 1.2e-37;
 Matches 114; Conservative 19; Mismatches 52; Indels 32; Gaps 5;
 QY 10 AGAENGSAAGGANGGAQAQAHPAIREQDRLMPTANVIRIMRRLVPAHAKISDDAKETI 69
 Db 47 asdsaaatgeene-----ctvreqdrfmpianvirimrkilpphakisddaketi 97
 QY 70 QECVSEYISFITGEANERCQQRKTTAEDVLWAMSRGLGFDYVEPLGAYLHRYREPEG 129
 Db 98 qecvseyisfitgeanercqrkitaedvlwamskigfdyieplumylhryreleg 157
 QY 130 DARGVGLVPGAAPSRGDHHPHSMSPAMKSRGPVSGAAMLPHHHHHHDMQMHAAWYGG 189
 Db 158 drtsm-----rgeplgkrtyeatl-----gvatafvpppyhhngyfgaampmg 202
 QY 190 T-----AVPPPAGPPHGGFLMPPHPOGSSHYLPYAYEP 222
 Db 203 tyvreappntasshhhhhhhhhhargisn-----ahep 235
 RESULT 5
 AAW71722
 ID AAW71722 standard; Protein; 208 AA.
 XX
 XX AAW71722;
 AC

Db 68 yisvtgeancrcqrearkitaeidilwamskigfdnyvqplvtvfnryreietd-rgsa 126
 QY 136 LVPGAAPSRGDHPHSMSPAAMLKSRGPVSGAAMLPHHHHHHDMQMHAAWYGTAVPPP 195
 Db 127 l-----rg-----eppslrqt-----ygnngl--- 143
 QY 196 AGPHHGGFLMPHQSSHYL-----PYAYPTYGGEHMAAYYGAAYAPGNGSGDGG 251
 Db 144 -----gf-----hpsghlpppgpyg-----gmlqsgmvggryyq-ngsssgqdes 185
 QY 252 SGGGGGSAS 260
 Db 186 svggsgss 194

RESULT 7
 ID AAY96222 standard; Protein; 280 AA.
 AC AAY96222;
 XX
 DT 11-SEP-2000 (first entry)
 DE Wheat LEC1.
 KW wheat; leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX
 OS Triticum aestivum.
 PN WO200028058-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-US26514.
 XX
 PR 09-NOV-1998; 98US-0107643.
 PR 10-NOV-1998; 98US-0107810.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX
 DR WPI: 2000-376568/32.
 DR N-PSDB; AAA27462.
 XX

PT New HAP3-type CCAAT-box binding transcriptional activators,
 PT particularly leafy cotyledon 1 transcriptional activator, useful for
 PT inducing somatic embryogenesis or apomixis in a plant cell -
 XX
 PS Claim 13; Page 93; 94pp; English.
 XX

CC The present sequence is the wheat leafy cotyledon 1 transcriptional
 CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
 CC LEC1 expression initiates the formation of embryo-like structures and
 CC improves growth and recovery of transformation by asexual reproduction, i.e.
 CC the replacement of sexual reproduction by asexual reproduction, LEC1
 CC expression in the nucellus integument, or cell specific expression in
 CC the megaspore mother cell would trigger embryo formation from maternal
 CC tissues only. This results in the production of seeds identical to the
 CC parent. Using LEC1, transgenic high yielding seeds could be developed.
 CC In addition, LEC1 could be used for positive selection of a transformed cell
 CC (transgenic plant), for increasing transformation efficiency and for
 CC increasing recovery of regenerated plants.
 XX
 SQ Sequence 280 AA;

Query Match 32.3%; Score 489.5; DB 21; Length 280;
 Best Local Similarity 41.5%; Pred. No. 6.3e-36;

Matches 119; Conservative 27; Mismatches 92; Indels 49; Gaps 9;
 QY 11 GAENSGAAGGANNNGAQAQAAPAIREDRLMPIANVIRMRVLPAAKISDDAKETIQ 70
 Db 5 gypngpaepaptqg-----tpvvregrlmpianvirrralpahakisddakeaiq 57
 QY 71 ECVSEVISPTGEANERCOEOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFE--- 128
 Db 58 ecvsefisvtgeanercrmqhrktvnaedivwalrfgdyvvpvlsvflhrmdpeag 117
 QY 129 -----GAARGVGLVPGAAPSRGSDHHPHSM-----SPAAMLKSRGPVSGAAMLPH 173
 Db 118 tggaaagdsrav-----tsappraappviahvplqagrmpyappaplvqenmqtrpyap- 172
 QY 174 HHHHHDMQMHAAWYGTAVPPPAGPPHHG---GFLMPHQSSSHYLYAYEPTYGGEHAM 230
 Db 173 -papvqvqmqrglygpra-----pvhgyavgmavpranvvggy-----qvfggegm 218
 QY 231 AAYYGAAYAPGNGSGDGGSGGGG---GSASHTPPQSGGLEHHPHP 274
 Db 219 aqyygygyeagaygagssnggaalgdeessngvvpagpgegmgep 265

RESULT 8
 ID AAY96217 standard; Protein; 214 AA.
 AC AAY96217;
 XX
 DT 11-SEP-2000 (first entry)
 DE Veronia mespilifolia LEC1.
 KW Leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
 KW selectable marker; transgenic plant; transgenic seed; HAP3.
 XX
 OS Veronia mespilifolia.
 PN WO200028058-A2.
 XX
 PD 18-MAY-2000.
 XX
 PF 09-NOV-1999; 99WO-US26514.
 XX
 PR 09-NOV-1998; 98US-0107643.
 PR 10-NOV-1998; 98US-0107810.
 XX
 PA (PION-) PIONEER HI-BRED INT INC.
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
 PI Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
 XX
 DR WPI: 2000-376568/32.
 DR N-PSDB; AAA27457.
 XX

PT New HAP3-type CCAAT-box binding transcriptional activators,
 PT particularly leafy cotyledon 1 transcriptional activator, useful for
 PT inducing somatic embryogenesis or apomixis in a plant cell -
 XX
 PS Claim 13; Page 83-84; 94pp; English.
 XX

CC The present sequence is the leafy cotyledon 1 transcriptional activator,
 CC LEC1 of Veronia mespilifolia. This sequence is a HAP3-type CCAAT-box
 CC binding protein. LEC1 expression initiates the formation of embryo-like
 CC structures and improves growth and recovery of transformation. During
 CC apomixis, i.e. the replacement of sexual reproduction by asexual
 CC reproduction, LEC1 expression in the nucellus integument, or cell
 CC specific expression in the megaspore mother cell triggers embryo
 CC formation from maternal tissues only. This results in the production of
 CC seeds identical to the parent. Using LEC1, transgenic high yielding seeds
 CC could be developed. In addition, LEC1 could be used for positive
 CC selection of a transformed cell (transgenic plant), for increasing

XX The present sequence is the soybean leafy cotyledon 1 transcriptional
CC activator, LEC1. This sequence is a HAP3-type CCAAT-box binding protein.
CC LEC1 expression initiates the formation of embryo-like structures and
CC improves growth and recovery of transformation. When apomixis occurs, i.e.
CC the replacement of sexual reproduction by asexual reproduction, LEC1
CC expression in the nucellus integument, or cell specific expression in
CC the megaspore mother cell would trigger embryo formation from maternal
CC tissues only. This results in the production of seeds identical to the
CC parent. Using LEC1, transgenic high yielding seeds could be developed. In
CC addition, LEC1 could be used for positive selection of a transformed cell
CC (transgenic plant), for increasing transformation efficiency and for
CC increasing recovery of regenerated plants.

SQ Sequence 171 AA;
Query Match 29.3%; Score 443; DB 21; Length 171;
Best Local Similarity 71.7%; Pred. No. 4.8e-32;
Matches 86; Conservative 7; Mismatches 27; Indels 0; Gaps 0;
QY 31 AAPAIREQRLMPLIANVIRIMRVLPAHAKISDDAKETIQECVSEYISFITGEANERCOR 90
DB 3 amagvtreqdympianvirimrilpahakisddaketigecvseyisfitaeancqr 62
QY 91 EQRKTTAEEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAAPSRGGDHP 150
DB 63 eqrktvtaedvlwameklgfnahplslhryresegepavrrassamginnnmvhp 122

RESULT 11
AAB33002
ID AAB33002 standard; Protein; 219 AA.
XX AAB33002;
XX 25-JAN-2001 (first entry)
XX Pinus radiata transcription factor protein sequence #129.
XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bz1p; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB.
OS Pinus radiata.

XX WO200053724-A2.
XX 14-SEP-2000.
XX 09-MAR-2000; 2000WO-US06112.
XX 11-MAR-1999; 99US-0266513.
XX 18-AUG-1999; 99US-0149485.
XX (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX Wood M, McGrath A, Shenk MA, Glenn M;
XX WPI; 2000-579369/54.
XX New isolated polynucleotide encoding a plant transcription factor for
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,
XX having modified gene expression or modified activity of a polypeptide
XX Claim 8; Page 387; 747pp; English.

XX The present invention relates to novel plant transcription factors from
XX Eucalyptus grandis or Pinus radiata. The present sequence is one such

CC transcription factor. The transcription factor may be used to produce a
CC plant having modified gene expression such as a woody plant e.g. a
CC eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or
CC to modify the activity of a polypeptide in a plant. The transcription
CC factors of the present invention are members from the following families
CC of regulatory proteins: bz1p, bz2p family of G-box binding factors, basic
CC helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain
CC zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2
CC Cys2His2, CCAAT box elements and MYB.
XX Sequence 219 AA;

Query Match 27.8%; Score 420.5; DB 21; Length 219;
Best Local Similarity 39.6%; Pred. No. 6.5e-30;
Matches 103; Conservative 29; Mismatches 71; Indels 57; Gaps 8;
QY 17 AAGGANNCGAAQQAAP-AIREQDRLMPLIANVIRIMRVLPAHAKISDDAKETIQECVSE 75
DB 4 apgrttnpaavtwaahsfvkeqdrfplanvgrlmkpalpangkvskdaketygecvse 63
QY 76 YISFITGEANERCORQKRTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGV 135
DB 64 fislfitgeasdkcqrkrktngdallwamtltfedyveplkiylhkyrmegekvsma 123
QY 136 LVPGAAPSRGGDHPHSHMS---PAAMLKSRGPVSG-----AAMLPHHHHHHDMQMAHY 187
DB 124 kqgdtpskgnnalngsslenpnaysglnpgynrvqsgslph-----mqgaay 175
QY 188 GGTAVPPPAGP--PHHGGFLMPHPQSSHYLPYAYEPTYGGEHAMAAYYGAAYAPNGG 245
DB 176 g---qppgmvyghg-----himgay-----nmt 197
QY 246 SGDSGSGSGSGSGSHTPPQG 265
DB 198 apnssggnssgqqqqqaprg 217

RESULT 12

AAY96225
ID AAY96225 standard; Protein; 90 AA.
XX AAY96225;
XX 11-SEP-2000 (first entry)
XX Arabidopsis LEC1 consensus region.
XX Leafy cotyledon 1 transcriptional activator; LEC1; apomixis;
XX selectable marker; transgenic plant; transgenic seed; HAP3.
XX Arabidopsis sp.
XX WO200028058-A2.
XX 18-MAY-2000.
XX 09-NOV-1999; 99WO-US26514.
XX 09-NOV-1998; 98US-0107643.
XX 10-NOV-1998; 98US-0107810.
XX (PION-) PIONEER HI-BRED INT INC.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX Lowe KS, Gordon-Kamm WJ, Klein TM, Rasco-Gaunt S, Cahoon RE;
XX Sun X, Hoerster GJ, Gregory CA, Nadimpalli R;
XX WPI; 2000-376568/32.
XX New HAP3-type CCAAT-box binding transcriptional activators,
XX particularly leafy cotyledon 1 transcriptional activator, useful for
XX inducing somatic embryogenesis or apomixis in a plant cell -

```

XX Example 6; Fig 1; 94pp; English.
PS
XX The present sequence is the leafy cotyledon 1 transcriptional activator,
XX LEC1 consensus sequence from Arabidopsis. This sequence is a HAP3-type
CC CCAAT-box binding protein. LEC1 expression initiates the formation of
CC embryo-like structures and improves growth and recovery of
CC transformants. When apomixis occurs, i.e. the replacement of sexual
CC reproduction by asexual reproduction, LEC1 expression in the nucellus
CC integument, or cell specific expression in the megaspore mother cell
CC would trigger embryo formation from maternal tissues only. This results
CC in the production of seeds identical to the parent. Using LEC1,
CC transgenic high yielding seeds could be developed. In addition, LEC1
CC could be used for positive selection of a transformed cell (transgenic
CC plant), for increasing transformation efficiency and for increasing
CC recovery of regenerated plants. The present sequence was used in the
CC identification of the plant LEC1 consensus sequence (AAY96223).
XX
XX Sequence 90 AA;
SQ
Query Match 27.0%; Score 409; DB 21; Length 90;
Best Local Similarity 83.3%; Pred. No. 2.4e-29;
Matches 75; Conservative 11; Mismatches 4; Indels 0; Gaps 0;
QY 36 REODRLMPIANVIRIMRRVLPAAKISDDAKETIQECVSEYISFITGEANERCQREQRT 95
DB 1 reqdymplanvirimrktlpshakisdsketiqecvseysfvtgeanercqreqrt 60
QY 96 ITAEDVLWMSRLGFDYDVEPLGAYLHYR 125
DB 61 itaedilwamsklgfdnyvqpltvfinryr 90
RESULT 13
AAG04651
ID AAG04651 standard; Protein: 141 AA.
XX
AC AAG04651;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 757.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 08-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.

```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150366.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 28-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

Query Match 24.9%; Score 376.5; DB 21; Length 141;
Best Local Similarity 46.0%; Pred. No. 3.2e-26;
Matches 75; Conservative 29; Mismatches 30; Indels 29; Gaps 3;

QY 8 PAAGAENGSAAGSANGGAAQQAAPATREODRLMPIANVIRMRVLPAAHAKISDDAKE 67
Db 5 pspagdggesgg-----svreqdryipianisrimkkaipngkigkdakd 51
QY 68 TIQECVSEYISITGEANRCOREQKKTITAEEDVLWMSRLGDDYVEPLGAYLHRYREF 127
Db 52 tvqecvsefifitaseaskckekrktvngddllwamatigfedyleplkiylaryrel 111
QY 128 EGDARGVGLVPGAAPSRGDDHHPHSMSPAAMLKSRGPVSGAAM 170
Db 112 egdnkgsg-----ksqd-----gsnrdaqgvgvsgeem 138

RESULT 14

AAG31928
ID AAG31928 standard; Protein; 138 AA.

XX AAG31928;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 38427.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 03-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135133.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139465.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0158458.
PR 29-SEP-1999; 99US-0158596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.


```

PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158223.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.

Query Match      23.8%; Score 360.5; DB 21; Length 139;
Best Local Similarity 45.4%; Pred. No. 8.4e-25;
Matches 74; Conservative 29; Mismatches 29; Indels 31; Gaps 4;
QY 8 PAAGAENGSAAGANNNGAAQQAAPAIHQDRLMPTIANVIRIMRVLPAHAKISDDAKE 67
Db 5 pspagdggesg-----svredrylpianisrlnkkalppngkigkdakd 51
QY 68 TIQECVSEYISFITGEANERCQREQRTITAEADVILWMSRLGDDYVEPLGAYLHRYREF 127
Db 52 tvqecvsefifitseasdkckrktvngddllwamatlgfedylepkiylaryr-- 109
QY 128 EGDARGVGLVPGAAPSRGGDHHPHSMSPAAMLKSRGPVSGNAAM 170
Db 110 egdnkgsg-----ksgd-----gsnrdaqggvgvsggeem 136

Search completed: October 25, 2001, 09:21:24
Job time: 284 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:20:20 ; Search time 16.39 seconds
(without alignments)
1292.039 Million cell updates/sec

Title: US-09-435-054-2
Perfect score: 1514
Sequence: 1 MDSSFLPAAGAENCSAAG.....ASHTPGSGGLEPHPPAYK 278
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	494	32.6	208	2 G86352	protein T26F17.20
2	410.5	27.1	161	2 G71407	transcription fact
3	410	27.1	228	2 T45874	transcription fact
4	391	25.8	179	2 S22820	transcription fact
5	372	24.6	178	2 A84788	probable CCAAT-box
6	360.5	23.8	138	2 E84810	hypothetical prote
7	357.5	23.6	160	2 G84919	probable CCAAT-box
8	354.5	23.4	215	2 F84508	probable CCAAT-box
9	324.5	21.4	205	2 S22817	transcription fact
10	324.5	21.4	207	2 F38245	transcription fact
11	324.5	21.4	207	2 A23692	transcription fact
12	322	21.3	180	2 S24469	transcription fact
13	320	21.1	186	2 JC6080	transcription fact
14	319.5	21.1	209	2 S22818	transcription fact
15	307	20.3	139	2 C86222	transcription fact
16	295.5	19.5	144	1 A28123	hypothetical prote
17	274	18.1	116	2 S42744	transcription fact
18	265	17.5	205	2 S51565	transcription fact
19	232.5	15.4	122	2 S22819	transcription fact
20	195	12.9	159	2 S53582	transcription fact
21	190.5	12.6	162	2 T50504	transcription fact
22	176.5	11.7	370	2 I37555	TAIR-binding prote
23	171	11.3	590	2 A26638	DRI-like protein -
24	170	11.2	369	2 TVFVAF	c-Maf protein - mo
25	161	10.6	176	2 A43320	homeotic protein D
26	160	10.6	175	2 JC5365	transforming prote
27	147.5	9.7	963	2 T48707	TATA-binding prote
28	147	9.7	1184	2 G01763	TBP-binding repres
29	147	9.7	1184	2 S50832	related to regulat
					atrophin-1 - human
					atrophin-1 - human

30 142 9.4 161 2 T40194
31 140.5 9.3 420 2 I59234
32 138 9.1 420 2 A49642
33 138 9.1 839 2 F75518
34 138 9.1 1212 2 T13804
35 138 9.1 1585 2 T31611
36 135.5 8.9 561 2 T16148
37 134 8.9 427 2 A23372
38 134 8.9 549 2 B23372
39 132.5 8.8 512 2 S70644
40 129.5 8.6 200 2 S10334
41 129 8.5 440 2 S71795
42 129 8.5 910 2 A34721
43 129 8.5 911 2 B34721
44 129 8.5 1166 2 T13958
45 129 8.5 1249 2 T14270

probable transcript
octamer binding tr
transcription fact
hypothetical prote
shs protein - frui
hypothetical prote
hypothetical prote
female-specific do
male-specific doub
annexin VII - Afri
glycine-rich prote
transcription fact
androgen receptor
androgen receptor
syngap-bl protein
Ras-GTPase activat

ALIGNMENTS

RESULT 1
G86352
protein T26F17.20 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86352
R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: G86352
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-208 <STO>
A:Cross-references: GB:AE005172; NID:g6552738; PIDN:AAF16537.1; GSPDB:GN00141
C:Genetics:
A:Gene: T26F17.20
A:Map position: 1

Query Match 32.6%; Score 494; DB 2; Length 208;
Best Local Similarity 47.0%; Pred. No. 3.2e-29;
Matches 117; Conservative 19; Mismatches 47; Indels 66; Gaps 9;

QY 16 SAAGGANNGAQAQAATREQRLMPIANVIRMRVLPANAKISDDAKETIQECVSE 75
DB 8 AGADKNGIVVQQPCVAREQQYMPIANVIRMRKTLPSHAKISDDAKETIQECVSE 67
QY 76 YISFITGEANERCQQRKTTAEDVLWMSRLGDDYVEPLGAYLHRYREFEGDARGVG 135
DB 68 YISFVTGEANERCQQRKTTAEDILWMSKLGFDNYVPLVFINRYREIETD-RGSA 126
QY 136 LVPGAAPSRGDDHPHSHMSPAAMLSKRGVSGAAMLPHHHHHDMQMAAAYGGTAVPPP 195
DB 127 L-----RG-----EPPSLRQT-----YGGNGI--- 143
QY 196 AGPHHGGLMHPHQQSSHYL----PVAYEPTYGGEHMAAAYGGRAYAPCNGSGDGG 251
DB 144 -----GF-----HGFSHGLPPPGPYG-----GMLDQSMVMGGGRYYO-NGSSGODES 185
QY 252 SGGGGSGAS 260
DB 186 SVGGSGSSS 194

4
RESULT
S22820
transcription factor NF-Y, CCAAT-binding, chain B - maize
N:Alternate names: CCAAT-box DNA-binding protein
C:Species: Zea mays (maize)
C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 23-Jul-1999
C:Accession: S22820
R:L.I., X.Y.; Mantovani, R.; van Huijsduijnen, R.H.; Andre, I.; Benoist, C.;
Nucleic Acids Res. 20, 1087-1091, 1992
A:Title: Evolutionary variation of the CCAAT-binding transcription factor
A:Reference number: S22816: MUID:92195809

Qy	8	PAAGA---ENGSAAGGANNGAQOHAAPAIROQDRMLPIANVIMRRLVPAHAKISDD	64
Db	8	PGGGGSHESGSPGGGGG-----SVRODRFLPIANIRKMAIPANGKIAKD	58
Qy	65	AKETIQCVSYISFITGEANNCERQQRKTTAEDVLWMSRLGDFDYVPLGAYLHY	120
Db	59	AKETIQCVSEFISFITSEADKQCRKRTIINGDILLWAMATLGFEDIYIPLKVYLOK	118
Qy	125	REFEGDARGVGLVPCAAPSRGGD-----HHPHSMSPAAMLKSRGPYS-GAAMLPHHH	170
Db	119	PMWFGSK-----LTAKSSDNGSTKKDAIGHVGAASSAAGGCGCAGYNOGMYMOPOH	170

Db	119	REMEGDSK-----LTAKSSDGSITKDALGHVGASSSAAEGMGQCGAYNQGMGYNQPYH	17
Qy	177	HHDM 180	
		: :	
Db	173	NGDI 176	

A;Residues: 1-178 <STO>
A;C/Cross-references: GB:AE002093; NID:gn371295; PIDN:AAD18153.1; GSPDB:GM00139
C;Genetics:
A;Gene: At2g37060

A:Map position: 2

Query Match 24.6%; Score 372; DB 2; Length 178;
Best Local Similarity 43.6%; Pred. No. 2.1e-20;
Matches 79; Conservative 27; Mismatches 49; Indels 26; Gaps 4;
QY 20 GANNNGAAQQAHPAIREQDRLPIANVIRIMRVLPAAKISDDAKETIOECVSEYISF 79
DB 13 GSHESGGDQSPRSLHVREODRLPIANVIRIMRVLPAAKISDDAKETIOECVSEYISF 79
QY 80 ITGEANERQOREKTTITAEVDLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPG 139
DB 73 VTSEADSKQREKRTKINGDLDLWAMATLGFEDYMEPLKYLHRYR-EGDTKG----- 124
QY 140 AAPSRGDGHHPHSMSPAAMLKSRGPVSGAAMLPHHHHDMQMHAAAMYGGTAVPPGAPPP 199
DB 125 --SAKGGD-----PNAKKDGOSSONG-----QFSQALAHQPGYNSQVTFPLESS 166
QY 200 H 200
DB 167 H 167
RESULT 6
E84810
hypothetical protein At2g38880 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: E84810
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: E84810
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-138 <STO>
A:Cross-references: GB:AE002093; NID:g3928076; PIDN:AAC79602.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g38880
A:Map position: 2

Query Match 23.8%; Score 360.5; DB 2; Length 138;
Best Local Similarity 45.4%; Pred. No. 1.1e-19;
Matches 74; Conservative 29; Mismatches 29; Indels 31; Gaps 4;
QY 8 PAAGBNNGSAGGANNNGAAQQAHPAIREQDRLPIANVIRIMRVLPAAKISDDAKE 67
DB 5 PSSPAGDGGESG-----SVREODRLPIANVIRIMRVLPAAKISDDAKE 51
QY 68 TIOECVSEYISFITGEANERQOREKTTITAEVDLWMSRLGFDYVEPLGAYLHRYREF 127
DB 52 TVQECVSEYISFITGEANERQOREKTTITAEVDLWMSRLGFDYVEPLGAYLHRYREF 127
QY 128 EGDARGVGLVPGAAAPSRGDDHHPHSMSPAAMLKSRGPVSGAAM 170
DB 110 EGDNKGSG-----KSGD-----GSRNDRAGGVSGEEM 136
RESULT 7
G84919
probable CCAAT-box binding transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84919
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: G84919
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <STO>
A:Cross-references: GB:AE002093; NID:g3738293; PIDN:AAC63635.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g47810
A:Map position: 2

Query Match 23.6%; Score 357.5; DB 2; Length 160;
Best Local Similarity 53.1%; Pred. No. 2.2e-19;
Matches 68; Conservative 23; Mismatches 26; Indels 11; Gaps 2;
QY 28 QOHAAPATREQDRLPIANVIRIMRVLPAAKISDDAKETIOECVSEYISFITGEANER 87
DB 42 QOESMMYKQEDRLPIANVIRIMRVLPAAKISDDAKETIOECVSEYISFITGEANER 87
QY 88 QOREKRTTITAEVDLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAAAPSRGDD 147
DB 102 CHKEARKTVNGDDICWAMNLGFDYAAQLKKYLHRYRVLEGE-----KPNHKGK 151
QY 148 HHPHSMSP 155
DB 152 GSPKS-SP 158
RESULT 8
F84508
probable CCAAT-box binding transcription factor [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84508
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84508
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-215 <STO>
A:Cross-references: GB:AE002093; NID:g4558662; PIDN:AAD22680.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g13570
A:Map position: 2

Query Match 23.4%; Score 354.5; DB 2; Length 215;
Best Local Similarity 37.8%; Pred. No. 4.9e-19;
Matches 79; Conservative 31; Mismatches 54; Indels 45; Gaps 6;
QY 8 PAAGAEN-GSAGGANNNGAAQQAHPAIREQDRLPIANVIRIMRVLPAAKISDDAK 66
DB 13 PGVAETNPGSPSSKTNNNNN-----KEQDRFLPIANVIRIMRVLPAAKISDDAK 65
QY 67 ETIOECVSEYISFITGEANERQOREKTTITAEVDLWMSRLGFDYVEPLGAYLHRYRE 126
DB 66 ETVOECVSEYISFITGEANERQOREKTTITAEVDLWMSRLGFDYVEPLGAYLHRYRE 126
QY 127 FEGDARGVGLVPGAAAPS-----RGDDHHPHSMSPAAMLKSRGPVSGAAMPLPHHH 175
DB 126 TEKEKVN-----SPKQQOORQQOQQIOQNNHNYOFQEQD---QNNNNMCSYTSIHHH 176
QY 176 HHHDMMHAAAMYGGTAVPPGAPPPHHGGF 204
DB 177 -----PSPFLPVDHQPF 188
RESULT 9

S22817

transcription factor NF-Y, CCAAT-binding, chain B - human
N:Alternate names: CAAT-box DNA-binding protein
C:Species: Homo sapiens (man)
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C:Accession: S22817
R:Li, X.Y.; Mantovani, R.; van Huijsduijnen, R.H.; Andre, I.; Benoist, C.; Mathis, D.
Nucleic Acids Res. 20, 1087-1091, 1992

A:Title: Evolutionary variation of the CCAAT-binding transcription factor NF-Y.

A:Reference number: S22816; MUID:92195809

A:Accession: S22817

A:Molecule type: mRNA

A:Residues: 1-205 <LII>

A:Cross-references: EMBL:X59710; NID:g35049; PIDN:CAA42230.1; PID:g35050

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:51-140/Domain: DNA binding #status predicted <DNA>

Query Match 21.4%; Score 324.5; DB 2; Length 205;

Best Local Similarity 56.6%; Pred. No. 7.2e-17;

Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQEQ 93

Db 49 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECSEFISFITSEASERCHQEK 108

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146

Db 109 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFREAMKGEKGIG---GAVTATDG 158

RESULT 10

F38245

transcription factor NF-Y, CCAAT-binding, chain B - mouse

C:Species: Mus musculus (house mouse)

C:Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 26-Aug-1999

C:Accession: F38245; S12045

R:Li, X.Y.; Hooft van Huijsduijnen, R.; Mantovani, R.; Benoist, C.; Mathis, D.

J. Biol. Chem. 267, 8984-8990, 1992

A:Title: Intron-exon organization of the NF-Y genes. Tissue-specific splicing modifies a

A:Reference number: A38245; MUID:92250488

A:Accession: F38245

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-207 <LIA>

A:Cross-references: GB:M86215

R:Hooft van Huijsduijnen, R.; Li, X.Y.; Black, D.; Matthes, H.; Benoist, C.; Mathis, D.

EMBO J. 9, 3119-3127, 1990

A:Title: Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP-1/CBF) subun

A:Reference number: S12044; MUID:91006004

A:Accession: S12045

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-207 <HOQ>

A:Cross-references: GB:X53362; PIDN:CAA39024.1; PID:g53363

C:Superfamily: transcription factor HAP3

C:Keywords: alternative splicing; DNA binding; transcription regulation

Query Match 21.4%; Score 324.5; DB 2; Length 207;

Best Local Similarity 56.6%; Pred. No. 7.3e-17;

Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQEQ 93

Db 51 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECSEFISFITSEASERCHQEK 110

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146

Db 111 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFREAMKGEKGIG---GAVTATDG 160

RESULT 11

A23692

transcription factor, CCAAT-binding, chain A1 - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 04-Oct-1991 #sequence_revision 04-Oct-1991 #text_change 26-Aug-1999

C:Accession: A23692

R:Vuorio, T.; Maity, S.N.; de Crombrughe, R.

J. Biol. Chem. 265, 22480-22486, 1990

A:Title: Purification and molecular cloning of the "A" chain of a rat heteromeric CCA

A:Reference number: A23692; MUID:91093096

A:Accession: A23692

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-207 <VUO>

A:Cross-references: GB:M5045; GB:J05701; NID:g203352; PIDN:AAA40887.1; PID:g203353

C:Superfamily: transcription factor HAP3

C:Keywords: alternative splicing; DNA binding; transcription regulation

F:53-142/Domain: DNA binding #status predicted <DNA>

Query Match 21.4%; Score 324.5; DB 2; Length 207;

Best Local Similarity 56.6%; Pred. No. 7.3e-17;

Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIRQDRLMPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQEQ 93

Db 51 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECSEFISFITSEASERCHQEK 110

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146

Db 111 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFREAMKGEKGIG---GAVSATDG 160

RESULT 12

S24469

transcription factor NF-Y, CCAAT-binding, chain B - chicken

C:Species: Gallus gallus (chicken)

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 26-Aug-1999

C:Accession: S24469

R:Benoist, C.

submitted to the EMBL Data Library, January 1992

A:Reference number: S24469

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-180 <BEN>

A:Cross-references: EMBL:X59713; NID:g63690; PIDN:CAA42233.1; PID:g63691

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:51-140/Domain: DNA binding #status predicted <DNA>

Query Match 21.3%; Score 322; DB 2; Length 180;

Best Local Similarity 59.8%; Pred. No. 9.5e-17;

Matches 61; Conservative 16; Mismatches 25; Indels 0; Gaps 0;

QY 34 AIRQDRLMPIANVIRMRVLPAAKISDDAKETIQECVSEYISFITGEANERCQEQ 93

Db 49 SFREQDIYLPANVARIMKNAIPQTGKIADAKCEVCQECSEFISFITSEASERCHQEK 108

QY 94 KTTAEDVLWMSRLGFDYVEPLGAYLHRYREFEGDARGVG 135

Db 109 KTINGEDILFAMSTLGFDSYVEPLKLYLQKFREAMKGEKGIG 150

RESULT 13

JC6080

transcription factor HAP3 - Emericella nidulans

C:Species: Emericella nidulans, Aspergillus nidulans

C:Date: 21-Jan-1997 #sequence_revision 21-Jan-1997 #text_change 26-Aug-1999

C:Accession: JC6080

R:Papagiannopoulos, P.; Andrianopoulos, A.; Sharp, J.A.; Davis, M.A.; Hynes, M.J.

Mol. Gen. Genet. 251, 412-421, 1996

A:Title: The hapC gene of *Aspergillus nidulans* is involved in the expression of CCAAT-CD
A:Reference number: JC6080; MUID:96285564

A:Accession: JC6080

A:Molecule type: mRNA

A:Residues: 1-186 <PAP>

A:Cross-references: GB:U35341; NID:g1017715; PIDN:AAC49411.1; PID:g1017716

C:Date: 02-Mar-2001 #sequence_revision 06-Feb-1998 #text_change 26-Aug-1999

C:Accession: S22818; S78116

R:Li, X.Y.; Mantovani, R.; van Huijsduijnen, R.H.; Andre, I.; Benoist, C.; Mathis, D.

Nucleic Acids Res. 20, 1087-1091, 1992

A:Title: Evolutionary variation of the CCAAT-binding transcription factor NF-Y.

A:Reference number: S22818; MUID:92195809

A:Accession: S22818

A:Molecule type: mRNA

A:Residues: 1-209 <LIX>

A:Cross-references: EMBL:X59712

R:Benoist, C.

submitted to the EMBL Data Library, January 1992

A:Reference number: S78116

A:Accession: S78116

A:Molecule type: mRNA

A:Residues: 1-110, R', 112-209 <BEN>

A:Cross-references: EMBL:X59712; NID:g64217; PIDN:CAA4232.1; PID:g64218

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:54-143/Domain: DNA binding #status predicted <DNA>

RESULT 15

C86222

Hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86222

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon

ausen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzita

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86222

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-139 <STO>

A:Cross-references: GB:AF005172; NID:g1922961; PIDN:AAB70405.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match

Best Local Similarity 20.3%; Score 307; DB 2; Length 139;

Matches 51; Conservative 26; Mismatches 15; Indels 0; Gaps 0;

QY 37 EODRLMPIANVIRMRVLPAAHAKISDDAKETIOECVSEYISFITGEANERCOREQRTI 96

DB 3 DEDRLPIANVGRMLKQILPSNAKISKAOTVOECATEFISFVTCASEKCHRENKTV 62

QY 97 TAEDVLWAMSRIGFDYVEPLGAYLHRYREFE 128

DB 63 NGDDIWWALSTGLDNYADAVGRHLHYREAE 94

Search completed: October 25, 2001, 09:23:40

Job time: 200 sec

Query Match

Best Local Similarity 21.1%; Score 320; DB 2; Length 186;

Matches 67; Conservative 25; Mismatches 29; Indels 4; Gaps 3;

QY 35 IREODRLMPIANVIRMRVLPAAHAKISDDAKETIOECVSEYISFITGEANERCOREQRTI 94

DB 41 VKEODRLPIANVIRMRVLPAAHAKISDDAKETIOECVSEYISFITGEANERCOREQRTI 94

QY 95 TITAEVLWAMSRIGFDYVEPLGAYLHRYREFE 153

DB 101 TVNGEDILFAMTSLGFENYAEALKIVLSKYRETQ-SARGEHQNPSPSSGYAGCD--PSAE 157

QY 154 SPRAAM 158

DB 158 LPAAL 162

RESULT 14

S22818

transcription factor NF-Y, CCAAT-binding, chain B - sea lamprey

N:Alternate names: CCAAT-box DNA-binding protein

C:Species: Petromyzon marinus (sea lamprey)

C:Date: 29-Jan-1998 #sequence_revision 06-Feb-1998 #text_change 26-Aug-1999

C:Accession: S22818; S78116

R:Li, X.Y.; Mantovani, R.; van Huijsduijnen, R.H.; Andre, I.; Benoist, C.; Mathis, D.

Nucleic Acids Res. 20, 1087-1091, 1992

A:Title: Evolutionary variation of the CCAAT-binding transcription factor NF-Y.

A:Reference number: S22818; MUID:92195809

A:Accession: S22818

A:Molecule type: mRNA

A:Residues: 1-209 <LIX>

A:Cross-references: EMBL:X59712

R:Benoist, C.

submitted to the EMBL Data Library, January 1992

A:Reference number: S78116

A:Accession: S78116

A:Molecule type: mRNA

A:Residues: 1-110, R', 112-209 <BEN>

A:Cross-references: EMBL:X59712; NID:g64217; PIDN:CAA4232.1; PID:g64218

C:Superfamily: transcription factor HAP3

C:Keywords: DNA binding; transcription regulation

F:54-143/Domain: DNA binding #status predicted <DNA>

Query Match

Best Local Similarity 21.1%; Score 319.5; DB 2; Length 209;

Matches 65; Conservative 18; Mismatches 38; Indels 3; Gaps 1;

QY 11 GAEGSLASGDHDESGSK---DPREQDIYLPANVIRMRVLPAAHAKISDDAKETIQ 70

DB 32 GAEGSLASGDHDESGSK---DPREQDIYLPANVIRMRVLPAAHAKISDDAKETIQ 70

QY 71 ECVSEYISFITGEANERCOREQRTITAEVLWAMSRIGFDYVEPLGAYLHRYREFE 130

DB 89 ECVSEYISFITGEANERCOREQRTITAEVLWAMSRIGFDYVEPLGAYLHRYREFE 130

QY 131 ARGV 134

DB 149 EKG 152

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 25, 2001, 09:21:25 ; Search time 22.41 Seconds
(without alignments)
1641.267 Million cell updates/sec

Title: US-09-435-054-2
Perfect score: 1514
Sequence: 1 MDSSFLPAGAENSAAG.....ASHTPGSGGLEHPHPAYK 278

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues
Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL16.*
1: sp-archaea.*
2: sp-bacteria.*
3: sp-fungi.*
4: sp-human.*
5: sp-invertebrate.*
6: sp-mammal.*
7: sp-mhc.*
8: sp-organelle.*
9: sp-phage.*
10: sp-plant.*
11: sp-rodent.*
12: sp-unclassified.*
13: sp-vertebrate.*
14: sp-virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	499.5	33.0	205	10	Q9FGJ0	Q9fgj0 arabidopsis
2	494	32.6	208	10	O81130	O81130 arabidopsis
3	494	32.6	208	10	Q9SFD8	Q9sfid8 arabidopsis
4	443.5	29.3	187	10	O23634	O23634 arabidopsis
5	443.5	29.3	190	10	Q9FGJ3	Q9fgj3 arabidopsis
6	410.5	27.1	161	10	O23310	O23310 arabidopsis
7	410	27.1	228	10	Q9LFI3	Q9lfi3 arabidopsis
8	376.5	24.9	141	10	O23633	O23633 arabidopsis
9	372	24.6	178	10	O9ZOC3	O9zqc3 arabidopsis
10	360.5	23.8	138	10	O9SLG0	O9slg0 arabidopsis
11	357.5	23.6	160	10	O82248	O82248 arabidopsis
12	354.5	23.4	215	10	Q9SIT9	Q9sit9 arabidopsis
13	344.5	22.8	215	3	O59848	O59848 aspergillus
14	331.5	21.9	206	13	O73744	O73744 xenopus lae
15	324.5	21.4	148	11	Q63091	Q63091 rattus norv
16	320	21.1	186	3	Q00735	Q00735 emericeella
17	313.5	20.7	242	5	O76256	O76256 schistosoma
18	307	20.3	139	10	O04027	O04027 arabidopsis
19	289.5	19.1	183	5	Q9VIS6	Q9vis6 drosophila

20	260.5	17.2	123	10	Q9FV58	Q9fv58 arabidopsis
21	260	17.2	403	5	O17286	O17286 caenorhabdi
22	191	12.6	373	4	Q9UP93	Q9up93 homo sapien
23	191	12.6	403	4	O75444	O75444 homo sapien
24	190.5	12.6	162	10	Q9LEY8	Q9ley8 arabidopsis
25	185	12.2	181	5	Q9NKA2	Q9nka2 drosophila
26	185	12.2	183	5	Q9VJQ5	Q9vjq5 drosophila
27	174	11.5	586	5	Q9UAL9	Q9ual9 drosophila
28	174	11.5	589	5	Q9VI48	Q9vi48 drosophila
29	170	11.2	369	13	Q92171	Q92171 gallus gall
30	160	10.6	175	13	O13068	O13068 xenopus lae
31	158.5	10.5	299	6	O97922	O97922 gorilla gor
32	158	10.4	2230	5	Q9VY40	Q9vy40 drosophila
33	148.5	9.8	1175	11	P70200	P70200 mus musculus
34	147.5	9.7	301	6	O97923	O97923 hylobates l
35	147.5	9.7	963	3	Q9P759	Q9p759 neurospora
36	147	9.7	1182	4	O9A495	O9a495 homo sapien
37	146	9.6	815	5	Q9V9W8	Q9v9w8 drosophila
38	145.5	9.6	1211	5	O24523	O24523 drosophila
39	145	9.6	145	11	O9JKP7	O9jkp7 mus musculus
40	145	9.6	291	6	O97928	O97928 pongo pygma
41	144.5	9.5	359	5	O9VJK0	O9vik0 drosophila
42	144.5	9.5	1190	4	O9S621	O9s621 homo sapien
43	144	9.5	147	4	Q9NRF9	Q9nrf9 homo sapien
44	144	9.5	147	4	Q9NR32	Q9nr32 homo sapien
45	142.5	9.4	313	6	O97927	O97927 pan paniscu

ALIGNMENTS

RESULT 1
Q9FGJ0 PRELIMINARY; PRT; 205 AA.
AC Q9FGJ0:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE SIMILARITY TO CCAAT-BOX-BINDING TRANSCRIPTION FACTOR.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025628; BAB09093.1;
SQ SEQUENCE 205 AA; 23022 MW; 6A15396AE49D1983B CRC64;

Query Match	33.0%	Score	499.5	DB	10	Length	205
Best Local Similarity	61.3%	Pred. No.	5.3e-33				
Matches	103	Conservative	16	Mismatches	24	Indels	25
Gaps	4						
QY 22	NNGGAQQAQHAAPAREQDKLMPANVIRIMRVLPAAKISDDAKETQECVSEYISFIT 81						
DB 18	SNGGEE-----CTVREQDRFMPANVIRIMRVLPAAKISDDAKETQECVSEYISFIT 73						
QY 82	GNANERCOREQRTTAEDVLWMSRLGFDVVEPLGALYHRYREFEGDAGVGLVPGAA 141						
DB 74	GNANERCOREQRTTAEDVLWMSRLGFDVVEPLGALYHRYREFEGDAGVGLVPGAA 141						
QY 142	PSRGGDHHPHSPAMALK-----SRGPVSGAAMLPHHHHHD 179						
DB 133	-----SMTNGLVYKPNGTWTEYCAVGPVGIHMAQHYHRQN 170						
RESULT 2							

081130	PRELIMINARY;	PRT;	208 AA.
ID	081130		
AC	081130;		
DT	01-NOV-1998 (TReMBUrel. 08, Created)		
DT	01-NOV-1998 (TReMBUrel. 08, Last sequence update)		
DT	01-MAR-2001 (TReMBUrel. 16, Last annotation update)		
DE	CCAAT-BOX BINDING FACTOR HAP3 HOMOLOG.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;		
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;		
OC	Brassicales; Brassicaceae; Arabidopsis.		
ON	NCBI_TaxID=3702;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=WS-0;		
RX	MEDLINE=98319234; PubMed=9657152;		
RA	Lotan T., Ocho M., Yee K., West M.A.L., Lo R., Kwong R.W.,		
RA	Yamagishi K., Fischer R.L., Goldberg R.B., Harada J.J.;		
RT	"Arabidopsis LEAFY CONVELEDON1 is sufficient to induce embryo		
RT	development in vegetative cells.";		
RL	Cell 93:11195-1205(1998).		
DR	EMBL; AF036684; AAC39488.1; -.		
DR	HSSP; P19267; 1B6W.		
DR	InterPro; IPR000166; -.		
DR	InterPro; IPR000947; -.		
DR	Pfam; PF00808; CBFD_NFYB_HMF; 1.		
DR	PRINTS; PR006115; CCAAT5SUBUNTA.		
DR	SEQUENCE 208 AA; 22693 MW;		
SD	4E2D249AE2525DDB CRC64;		

Query Match 32.6%; Score 494; DB 10; Length 208;
Best Local Similarity 47.0%; Pred. NO. 1.5e-32;
Matches 117; Conservative 19; Mismatches 47; Indels 66; Gaps 9;

Qy	16	SAAGANNNGAAQQAAPATREQDRLMPITANIRIMRRVLPAAHAKISDDAKETIQECVSE	75
		: :	
Db	8	AGAGDKNNGIVVOQPPCVAREDDQYMPITANIRIMRKLTPSHAKISDDAKETIQECVSE	67
Qy	76	YISFTITGEANERCOREQKTTIADVLWAMSLGFDXVEPLGAYLHURYFEFGDARGVG	135
Db	68	YISFVTGEANERCOREQKTTIADILWAMSLGFDNYDPLTFVFNRYREIETD-RGSA	126
Qy	136	LVPGAAPSRGGDHHPHSMSPAAMLKSRGPVSGAAMLPHHHHHDMQMHAAMYGGTAVPPP	195
Db	127	L-----RG-----EPPSLRQT-----YGGNGI---143	
Qy	196	AGPPHHGGFLMPHPQGSNNL---PYAYEPTYGGEHAAAYGGAAYAPCNCGSGDGS	251
Db	144	-----CF-----HGSHGLPPPGYGY-----GMLDQSMVMGGRYYO-NGSSGQDES	185
Qy	252	SGGGGGSAS	260
Db	186	SYGGGSSS	194

RESULT	3
Q9SFD8	
ID	PRELIMINARY; PRT; 208 AA.
AC	Q9SFD8;
DT	01-MAY-2000 (TrEMBLrel. 13, Created)
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE	T26F17.20.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC	Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_taxID=3702;
RN	[1]

RA Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Khan S., Kim C., Altafi H., Bel Q., Chin C., Chiou J., Choi E., Conn L., Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,

RA	Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
RA	Palim C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,
RA	Toriuni M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA	Ecker J.R.;
RT	"Genomic sequence for Arabidopsis thaliana BAC T26F17 from chromosome
RT	I."
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AC013482; AAF16537.1; "-.
DR	HSSP; PI9267; 1B6W.
DR	InterPro; IPR000166; "-.
DR	InterPro; IPR000947; "-.
DR	Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR	PRINTS; PR00615; CCAATSUBUNTA.
SQ	SEQUENCE 208 AA; 22679 MW; 0DEB469D9F8BFDFC CRC64;

Query Match	32.6%;	Score 494;	DB 10;	Length 208;
Best Local Similarity	47.0%;	Pred. No. 1.5e-32;		
Matches 117;	Conservative	19;	Mismatches 47;	Indels 66;
				Gaps 9;

Qy	16	SAGAGANNNGGAQQHAPATREODRLMPIANVIRIMRRVLPAHAKTSSDDAKETIQECVSE	75
		: : : : : :	
Db	8	AGAGDKNNGLVVQQCPPCVAREQDQYMPIANVIRIMRKLPSIAKTSDDAKETIQECVSE	67
		: : : : : :	
Qy	76	YISFTIGEANERCQRQRKITTAEDVLWAMSLRGFDYVEPLCAYLHRYREREGDARGVG	135
		: : : : : :	
Db	68	YISFVTGEANERCQRQRKITTAEDILWAMSKLGFDNYVDPTVFIVNRIRETLD-RGSA	126
		: : : : : :	
Qy	136	LVPGAAPSRRGGDHHPHSMSPAAMLKSRGPVSGAAMLPHHHHHMDQMIAAMYGGTAVPPP	195
		: : : : : :	
Db	127	L-----RC-----EPSSLRQT-----YGNGI---	143
Qy	196	AGPPHHGGFLMHPQGSSHYL----PYAVEPTYGGEIHAAAYGGNAYPGNCGSGDSGS	251
		: : : : : :	
Db	144	-----GF-----HGSHGLPPPGTYG-----GMLDSMWVGGGRYYIQ-NGSSSQDES	185
Qy	252	SGGGGGSAS	260
		: : : : : :	
Db	186	SVGGGSSSS	194

```

RESULT      4
O23634
ID          O23634      PRELIMINARY;      PRT;      187 AA.
AC          O23634;
DT          DT
DT          01-JAN-1998 (TrEMBLrel. 05, Created)
DT          01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT          01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE          TRANSCRIPTION FACTOR (FRAGMENT).
DE          HAP3B.
OS          Arabidopsis thaliana (Mouse-ear cress).
OC          Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC          Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC          Brassicales; Brassicaceae; Arabidopsids.
OX          NCBI_Taxid=3702;
RN          [1]
RP          SEQUENCE FROM N.A.
RA          Edwards D., Smith A.G., Murray J.A.;
RL          Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR          EMBL: Y13724; CAA74052.1; -.
DR          Mendel; 26811; Atath;2691;26811.
DR          InterPro: IPR000166; -.
DR          InterPro: IPR000947; -.
DR          Pfam: PF00808; CBFD.NFYB.HMF. 1.
DR          PRINTS: PR00615; CCAATSUBUNTA.
DR          PROSITE: PS00685; CBFA_NFYB; 1.
FT          NON_TER      1
SQ          SEQUENCE      187 AA; 20295 MW; E4C7ABC89F37F40E CRC64.

```

Query Match 29.3%; Score 443.5; DB 10; Length 187;
Best Local Similarity 38.7%; Pred. No. 1.6e-28;
Matches 94; Conservative 29; Mismatches 55; Indels 65; Caps 3;


```

RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; ALI32966; CAB67641.1; -
DR InterPro; IPR000166; -
DR InterPro; IPR000911; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR ProDom; PD001367; -; 1.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 228 AA; 25132 MW; BC158AE0A08579B6 CRC64;

Query Match 27.1%; Score 410; DB 10; Length 228;
Best Local Similarity 49.7%; Pred. No. 1e-25;
Matches 78; Conservative 29; Mismatches 48; Indels 2; Gaps 1;

QY 10 AGAENGSAAGGANNCAOQHAAPATREODRLMPIANVIRIMRRLPAHAKISDDAKETI 69
DB 10 AGAENGSAAGGANNCAOQHAAPATREODRLMPIANVIRIMRRLPAHAKISDDAKETI 69
QY 2 AEOQTGGGGGGHSGGDSQPSNLNREQDRFLPIANISIRIMKRGGLPLNGKIAKDAK 61
DB 2 AEOQTGGGGGGHSGGDSQPSNLNREQDRFLPIANISIRIMKRGGLPLNGKIAKDAK 61
QY 70 QECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFE 129
DB 70 QECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFE 129
QY 62 QECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFE 121
DB 62 QECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFE 121
QY 130 DARGCVGLVGAAPSRGGDHIHPSMSPAAMKSRGPVS 166
DB 130 DARGCVGLVGAAPSRGGDHIHPSMSPAAMKSRGPVS 166
QY 122 DTRKSG--KGESSAKRDGQPSQVFSQVPOQGSFS 156
DB 122 DTRKSG--KGESSAKRDGQPSQVFSQVPOQGSFS 156

RESULT 8
ID O23633 PRELIMINARY; PRT; 141 AA.
AC O23633;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE TRANSCRIPTION FACTOR.
GN HAP3A.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Edwards D., Smith A.G., Murray J.A.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y13723; CAA74051.1; -
DR HSP; P19267; 1B6W
DR Mendel; 26810; Arath; 2691; 26810.
DR InterPro; IPR000166; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 141 AA; 15181 MW; D33060B5AD21D9E0 CRC64;

Query Match 24.9%; Score 376.5; DB 10; Length 141;
Best Local Similarity 46.0%; Pred. No. 3e-23;
Matches 75; Conservative 29; Mismatches 30; Indels 29; Gaps 3;

QY 8 PAANGENSAAGGANNCAOQHAAPATREODRLMPIANVIRIMRRLPAHAKISDDAK 67
DB 8 PAANGENSAAGGANNCAOQHAAPATREODRLMPIANVIRIMRRLPAHAKISDDAK 67
QY 5 PSSPAGDGGGSG-----SVREQDRVLPIANISIRIMKRGGLPLNGKIAKDAK 51
DB 5 PSSPAGDGGGSG-----SVREQDRVLPIANISIRIMKRGGLPLNGKIAKDAK 51
QY 68 TIOECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREF 127
DB 68 TIOECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREF 127
QY 52 TVQECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREF 111
DB 52 TVQECVSEYISFITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREF 111
QY 128 EGDARGCVGLVGAAPSRGGDHIHPSMSPAAMKSRGPVSGAAM 170
DB 128 EGDARGCVGLVGAAPSRGGDHIHPSMSPAAMKSRGPVSGAAM 170

```

```

DB 112 EGDNRKSG-----KSGD-----GSNRDAGGGVSGEEM 138
DB 112 EGDNRKSG-----KSGD-----GSNRDAGGGVSGEEM 138

RESULT 9
QY2QC3
ID Q92QC3 PRELIMINARY; PRT; 178 AA.
AC Q92QC3;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE PUTATIVE CCAAT-BINDING TRANSCRIPTION FACTOR.
GN T2N18.18.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.,
RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Niernan W.C.,
RA Fraser C.M., Venter J.C.;
RA "Arabidopsis thaliana chromosome II BAC T2N18 genomic sequence.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006260; AAD18153.1; -
DR Mendel; 39697; Arath; 2691; 39697.
DR InterPro; IPR000166; -
DR InterPro; IPR000947; -
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 178 AA; 19523 MW; B3CF497383EC86C4 CRC64;

Query Match 24.6%; Score 372; DB 10; Length 178;
Best Local Similarity 43.6%; Pred. No. 9e-23;
Matches 79; Conservative 27; Mismatches 49; Indels 26; Gaps 4;

QY 20 GANNGAAQOHAAPATREODRLMPIANVIRIMRRLPAHAKISDDAKETIOECVSEYISF 79
DB 20 GANNGAAQOHAAPATREODRLMPIANVIRIMRRLPAHAKISDDAKETIOECVSEYISF 79
QY 13 GSHSGGDSQPSNLNREQDRFLPIANISIRIMKRGGLPLNGKIAKDAKIEIVQECVSEFISF 72
DB 13 GSHSGGDSQPSNLNREQDRFLPIANISIRIMKRGGLPLNGKIAKDAKIEIVQECVSEFISF 72
QY 80 ITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFEEDARGVGLVPG 139
DB 80 ITGEANERCOREOKTITAEEDVLWMSRLGFDYVEPLGAYLHRYREFEEDARGVGLVPG 139
QY 73 VTSEASDKCOREKRTKTINGDDLLWAMATLGFEDTMEPLKYLMYR--EGDTKG----- 124
DB 73 VTSEASDKCOREKRTKTINGDDLLWAMATLGFEDTMEPLKYLMYR--EGDTKG----- 124
QY 140 AAPSRGGDHIHPSMSPAAMKSRGPVSGAAMLPHHHHHDMOMHAMMYGCTAVPPAGPP 199
DB 140 AAPSRGGDHIHPSMSPAAMKSRGPVSGAAMLPHHHHHDMOMHAMMYGCTAVPPAGPP 199
QY 125 --SAKGGD-----PNAKKDGGSSQNG-----QFSQLAHQGPYGNSTVTFPLFS 166
DB 125 --SAKGGD-----PNAKKDGGSSQNG-----QFSQLAHQGPYGNSTVTFPLFS 166
QY 200 H 200
DB 167 H 167

RESULT 10
QYSLGO
ID Q9SLGO PRELIMINARY; PRT; 138 AA.
AC Q9SLGO;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE PUTATIVE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT.
GN AT2G38880.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]

```

RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
Salzhager G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC005770; AAC79602.1; -.
DR HSSP: P19267; 1B6W.
DR InterPro: IPR000166; -.
DR Pfam: PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PROSITE: PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 138 AA; 14764 MW; 14CE11FC21D5768A CRC64;

Query Match 23.4%; Score 360.5; DB 10; Length 138;
Best Local Similarity 45.4%; Pred. No. 5.7e-22;
Matches 74; Conservative 29; Mismatches 29; Indels 31; Gaps 4;

QY 8 PAAGAENGSAAGGANNAGAAQAAPAIREDQRLMPTIANVIRIMRRVLPAAHAKISDDAKE 67
Db 5 PSSPAGDGGESG-----SVREQDRYLPANISIRIMKALPPNGKICKGDKAD 51

QY 68 TIOECVSEYISFTIGANERCQEQRTTAEVDVWMSRLGFDVVEPLGAYLHRYREF 127
Db 52 TVQECVSEYISFTSEASDKQCKEKRKTNGDOLLWAMATLGPEDYLEPLKIYLYR-- 109

QY 128 EGDARGVGLVPGAAPRGDGHHPHSPAMLSKRGPSVSAAM 170
Db 110 EGONKSG-----KSGD-----GSNRDAGGGVSGEEM 136

RESULT 11
ID 082248 PRELIMINARY; PRT; 160 AA.
AC 082248;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMELrel. 16, Last annotation update)
DE PUTATIVE CCAAT-BINDING TRANSCRIPTION FACTOR SUBUNIT A (CBF-A).
GN F17A22.20.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
Salzberg S.L., Preuss D., Nierman W.C., White O., Eisen J.A.,
Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome II of Arabidopsis thaliana.";
RL Nature 402:761-768(1999).
DR EMBL: AC007063; AAD22680.1; -.
DR HSSP: P48781; 1B67.
DR InterPro: IPR000166; -.
DR InterPro: IPR000947; -.
DR Pfam: PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS: PR00615; CCAATSUBUNTA.
DR PROSITE: PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 215 AA; 24619 MW; 0615B2B7A097DD25 CRC64;

Query Match 23.4%; Score 354.5; DB 10; Length 215;
Best Local Similarity 37.8%; Pred. No. 2.9e-21;
Matches 79; Conservative 31; Mismatches 54; Indels 45; Gaps 6;

QY 8 PAAGAENGSAAGGANNAGAAQAAPAIREDQRLMPTIANVIRIMRRVLPAAHAKISDDAK 66
Db 13 PGVAETNPGSPSSKTNNNNN-----KEODRFLPIANVGRIMKKVLPNGKISKDAK 65

QY 67 ETIOECVSEYISFTIGANERCQEQRTTAEVDVWMSRLGFDVVEPLGAYLHRYRE 126
Db 66 ETVQECVSEYISFTIGANERCQEQRTTAEVDVWMSRLGFDVVEPLGAYLHRYRE 125

QY 127 FEGDARGVGLVPGAAPS-----RGDHPHSPAMLSKRGPSVSAAMLPHHH 175
Db 126 TEGEKVN-----SPKQQQQQQQQQQQQQHHYQEQED--QNNNNMSTYSISHHH 176

QY 176 HHMDMOMHAAMYGGTAVPPPPAGPHHGGF 204
Db 177 -----PSPFLPVHQPF 188

RESULT 13

Query Match 23.6%; Score 357.5; DB 10; Length 160;
Best Local Similarity 53.1%; Pred. No. 1.2e-21;
Matches 68; Conservative 23; Mismatches 26; Indels 11; Gaps 2;

```

059848
ID O59848 PRELIMINARY; PRT; 215 AA.
AC O59848;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HAPC.
GN HAPC.
OS Aspergillus oryzae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eukaryotes; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5062;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM02238;
RA Kato M., Kobayashi T., Tsukagoshi N., Tanaka A., Hashimoto H.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010431; BAA28356.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR000166;
DR InterPro; IPR000947;
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 215 AA; 23348 MW; 7A1A74E8F2E05C06 CRC64;

Query Match 22.8%; Score 344.5; DB 3; Length 215;
Best Local Similarity 38.5%; Pred. No. 1.9e-20;
Matches 77; Conservative 28; Mismatches 48; Indels 47; Gaps 4;

QY 27 AQHAAPATREQRLPIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANE 86
DB 33 AQOGEFEVKEQRLPIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANE 92
QY 87 RCQREKRTTADVLWMSRLGDDYVEPLGAYLHRYREFE----- 128
DB 93 KCOEKRTKNGEDILFAMTSLGFGFYAALYLSKYRETOSAGEHQNRRTSGYASG 152
QY 129 GDARGVGLVPGAPSRG-----DHHPHSMSPAAMLKSRGPVSGAALPHHHHHMDQM 182
DB 153 GPVGVSAPGGRPATAGFPDAADNTNSIMPSLDPTQDP----- 194
QY 183 HAAMYGSTAVPPGAPPHHG 202
DB 195 --SAYG---YPPMVGGPHNG 209

RESULT 14
ID O73744 PRELIMINARY; PRT; 206 AA.
AC O73744;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE NUCLEAR Y/CCAAT-BOX BINDING FACTOR B SUBUNIT NF-YB.
OS Xenopus laevis (african clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Herrier M., Wolffe A.P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF041204; AAC82336.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR000166;
DR InterPro; IPR000947;
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 206 AA; 22562 MW; 35933B20A6D43607 CRC64;

```

```

Query Match 21.9%; Score 331.5; DB 13; Length 206;
Best Local Similarity 51.9%; Pred. No. 2e-19;
Matches 69; Conservative 19; Mismatches 42; Indels 3; Gaps 2;

QY 34 AIREQRLMPTIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREQ 93
DB 50 SFREQDILPIANVARIMKNVAPQTKIAKDAKEQCVSEFISFITSEASERCHOEKR 109
QY 94 KTIITAEVDLWMSRLGDDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG--HHPH 151
DB 110 KTINGEDILFAMSLGFDYVEPLKLYLQKPFREAMKKGIGGTVTTGGDGLGEDLAEPF 169
QY 152 -SMSPAAMLKSRG 163
DB 170 TSQIPAGLITTDG 182

```

```

RESULT 15
ID O63091 PRELIMINARY; PRT; 148 AA.
AC O63091;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE CCAAT BINDING TRANSCRIPTION FACTOR-B SUBUNIT.
GN CBF-A11.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=91093096; PubMed=2266139;
RT "Purification and molecular cloning of the 'A' chain of a rat
RT heteromeric CCAAT-binding protein. Sequence identity with the yeast
RT HAP3 transcription factor."
RL J. Biol. Chem. 265:22480-22486(1990).
DR EMBL; M0617; AAA40888.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR000166;
DR InterPro; IPR000947;
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
DR PRINTS; PR00615; CCAATSUBUNTA.
DR PROSITE; PS00685; CBFA_NFYB; 1.
SQ SEQUENCE 148 AA; 16302 MW; 6855E5C857115AD0 CRC64;

```

```

Query Match 21.4%; Score 324.5; DB 11; Length 148;
Best Local Similarity 56.6%; Pred. No. 5e-19;
Matches 64; Conservative 17; Mismatches 29; Indels 3; Gaps 1;

QY 34 AIREQRLMPTIANVIRIMRRVLPAAHAKISDDAKETIQECVSEYISFITGEANERCQREQ 93
DB 13 SFREQDILPIANVARIMKNVAPQTKIAKDAKEQCVSEFISFITSEASERCHOEKR 72
QY 94 KTIITAEVDLWMSRLGDDYVEPLGAYLHRYREFEGDARGVGLVPGAPSRGG 146
DB 73 KTINGEDILFAMSLGFDYVEPLKLYLQKPFREAMKKGIG--GAVSATDG 122

```

Search completed: October 25, 2001, 09:24:34
Job time: 189 sec



```

EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est33.*
26: gb_est34.*
27: gb_est35.*
28: gb_est36.*
29: gb_est37.*
30: gb_est38.*
31: gb_est39.*
32: gb_est40.*
33: em_estba.*
34: em_estfun.*
35: em_esthum1.*
36: em_esthum2.*
37: em_esthum3.*
38: em_esthum4.*
39: em_esthum5.*
40: em_esthum6.*
41: em_esthum7.*
42: em_esthum8.*
43: em_esthum9.*
73: em_estp1.*
74: em_estp12.*
75: em_estp13.*
76: em_estp14.*
77: em_estp15.*
78: em_estp16.*
79: em_estp17.*
80: em_estp18.*
81: em_estp19.*
82: em_estp110.*
83: em_estro1.*
84: em_estro2.*
85: em_estro3.*
86: em_estro4.*
87: em_estro5.*
88: em_estro6.*
89: em_estro7.*
90: em_estro8.*
91: em_estro9.*
92: em_estro10.*
93: em_estro11.*
94: em_estro12.*
95: em_estro13.*
96: em_estro14.*
97: em_estro15.*
98: em_estro16.*
99: em_estro17.*
100: em_estro18.*
101: em_estro19.*
102: em_estro20.*
103: gb_est25.*
104: gb_est26.*
105: gb_est27.*
106: gb_est28.*
107: gb_est29.*
108: gb_est30.*
109: gb_est31.*
110: gb_est32.*
111: gb_est41.*
112: gb_est42.*
113: gb_est43.*
114: gb_est44.*
115: gb_est45.*
116: gb_est46.*
117: gb_est47.*

```

117:	gb_est48:	*
118:	gb_est49:	*
119:	gb_est50:	*
120:	gb_est51:	*
121:	gb_est52:	*
122:	gb_est53:	*
123:	gb_est54:	*
124:	gb_est55:	*
125:	gb_est56:	*
126:	gb_est57:	*
127:	gb_est58:	*
128:	gb_est59:	*
129:	gb_est60:	*
130:	gb_est61:	*
131:	gb_est62:	*
132:	gb_est63:	*
133:	gb_est64:	*
134:	gb_est65:	*
135:	gb_est66:	*
136:	gb_est67:	*
137:	gb_est68:	*
138:	gb_est69:	*
139:	gb_est70:	*
140:	gb_est71:	*
141:	gb_est72:	*
142:	gb_est73:	*
143:	gb_est74:	*
144:	gb_est75:	*
145:	gb_est76:	*
146:	gb_est77:	*
147:	gb_est78:	*
148:	gb_est79:	*
149:	gb_est80:	*
150:	gb_est81:	*
151:	gb_est82:	*
152:	gb_est83:	*
153:	gb_est84:	*
154:	gb_est85:	*
155:	gb_est86:	*
156:	gb_est87:	*
157:	gb_est88:	*
158:	gb_est89:	*
159:	gb_est90:	*
160:	gb_est91:	*
161:	gb_est92:	*
162:	gb_est93:	*
163:	gb_est94:	*
164:	gb_est95:	*
165:	gb_est96:	*
166:	gb_est97:	*
167:	gb_est98:	*
168:	gb_est99:	*
169:	gb_est100:	*
170:	gb_est101:	*
171:	gb_est102:	*
172:	gb_est103:	*
173:	gb_est104:	*
174:	gb_est105:	*
175:	gb_est106:	*
176:	gb_est107:	*
177:	gb_est108:	*
178:	gb_est109:	*
179:	gb_est110:	*
180:	gb_est111:	*
181:	gb_est112:	*
182:	gb_est113:	*
183:	gb_est114:	*
184:	gb_est115:	*
185:	gb_est116:	*
186:	gb_est117:	*
187:	gb_est118:	*
188:	gb_est119:	*
189:	gb_est120:	*
190:	gb_est121:	*
191:	gb_est122:	*
192:	gb_est123:	*
193:	gb_est124:	*
194:	gb_est125:	*
195:	gb_est126:	*
196:	gb_est127:	*
197:	gb_est128:	*
198:	gb_est129:	*
199:	gb_est130:	*
200:	gb_est131:	*
201:	gb_est132:	*
202:	gb_est133:	*
203:	gb_est134:	*
204:	gb_est135:	*
205:	gb_est136:	*
206:	gb_est137:	*
207:	gb_est138:	*
208:	gb_est139:	*
209:	gb_est140:	*
210:	gb_est141:	*
211:	gb_est142:	*
212:	gb_est143:	*
213:	gb_est144:	*
214:	gb_est145:	*
215:	gb_est146:	*
216:	gb_est147:	*
217:	gb_est148:	*
218:	gb_est149:	*
219:	gb_est150:	*
220:	gb_est151:	*
221:	gb_est152:	*
222:	gb_est153:	*
223:	gb_est154:	*
224:	gb_est155:	*
225:	gb_est156:	*
226:	gb_est157:	*
227:	gb_est158:	*
228:	gb_est159:	*
229:	gb_est160:	*
230:	gb_est161:	*
231:	gb_est162:	*
232:	gb_est163:	*
233:	gb_est164:	*
234:	gb_est165:	*
235:	gb_est166:	*
236:	gb_est167:	*
237:	gb_est168:	*
238:	gb_est169:	*
239:	gb_est170:	*
240:	gb_est171:	*
241:	gb_est172:	*
242:	gb_est173:	*
243:	gb_est174:	*
244:	gb_est175:	*
245:	gb_est176:	*
246:	gb_est177:	*
247:	gb_est178:	*
248:	gb_est179:	*
249:	gb_est180:	*
250:	gb_est181:	*
251:	gb_est182:	*
252:	gb_est183:	*
253:	gb_est184:	*
254:	gb_est185:	*
255:	gb_est186:	*
256:	gb_est187:	*
257:	gb_est188:	*
258:	gb_est189:	*
259:	gb_est190:	*
260:	gb_est191:	*
261:	gb_est192:	*
262:	gb_est193:	*
263:	gb_est194:	*
264:	gb_est195:	*
265:	gb_est196:	*
266:	gb_est197:	*
267:	gb_est198:	*
268:	gb_est199:	*
269:	gb_est200:	*
270:	gb_est201:	*
271:	gb_est202:	*
272:	gb_est203:	*
273:	gb_est204:	*
274:	gb_est205:	*
275:	gb_est206:	*
276:		

190:	gb_estll0.*
191:	gb_estll1.*
192:	gb_hc.*
193:	em_gss_fun.*
194:	em_gss_hum1.*
195:	em_gss_hum2.*
196:	em_gss_hum3.*
197:	em_gss_hum4.*
198:	em_gss_hum5.*
199:	em_gss_hum6.*
200:	em_gss_hum7.*
201:	em_gss_hum8.*
202:	em_gss_hum9.*
203:	em_gss_inv1.*
204:	em_gss_inv2.*
205:	em_gss_inv3.*
206:	em_gss_other.*
207:	em_gss_pln1.*
208:	em_gss_pln2.*
209:	em_gss_pro.*
210:	em_gss_rodl.*
211:	em_gss_rod2.*
212:	em_gss_rod3.*
213:	em_gss_rod4.*
214:	em_gss_rod5.*
215:	em_gss_vrt1.*
216:	em_gss_vrt2.*
217:	em_gss_vrt3.*
218:	gb_gss1.*
219:	gb_gss2.*
220:	gb_gss3.*
221:	gb_gss4.*
222:	gb_gss5.*
223:	gb_gss6.*
224:	gb_gss7.*
225:	gb_gss8.*
226:	gb_gss9.*
227:	gb_gss10.*
228:	gb_gss11.*
229:	gb_gss12.*
230:	gb_gss13.*
231:	gb_gss14.*
232:	gb_gss15.*
233:	gb_gss16.*
234:	gb_gss17.*
235:	gb_gss18.*
236:	gb_gss19.*
237:	gb_gss20.*
238:	gb_gss21.*
239:	gb_gss22.*
240:	gb_gss23.*
241:	gb_gss24.*
242:	gb_gss25.*
243:	gb_gss26.*
244:	gb_gss27.*
245:	gb_gss28.*
246:	gb_gss29.*
247:	gb_gss30.*
248:	gb_gss31.*
249:	gb_gss32.*
250:	gb_gss33.*
251:	gb_gss34.*
252:	em_gss_inv4.*
253:	em_gss_inv5.*
254:	em_gss_inv6.*
255:	em_gss_inv7.*
256:	em_gss_inv8.*
257:	gb_gss35.*
258:	gb_gss36.*
259:	gb_gss37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

BF595304

DB 79 TGAAGGTGTTACCTGCAGCGTTCAGGAGATGGAGGAGAAGCGTGCGCGCGTG 20
 QY 478 tcccggggcccgcacat 495
 DB 19 ATAGGAGCGTCCCTCCT 2

RESULT 10
 BF263449 595 bp mRNA EST 09-MAR-2001
 LOCUS HV.C2a006M10f Hordeum vulgare seedling green leaf EST library
 DEFINITION HVC2A006M10f (Erysiphe infected & control) Hordeum vulgare CDNA clone
 HV.C2a006M10f, mRNA sequence.

ACCESSION BF263449
 VERSION BF263449.2 GI:13260832
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
 REFERENCE 1 (bases 1 to 595)
 AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,
 Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
 ,T., Sanki,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
 Wood,T.
 TITLE Development of a genetically and physically anchored EST resource
 for barley genomics
 JOURNAL Unpublished (2000)
 COMMENT On Nov 17, 2000 this sequence version replaced gi:11194443.
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
 Seq primer: AATTAAACCCTCACTAAGGS
 High quality sequence cut: 594.
 Location/Qualifiers
 1..595
 /organism="Hordeum vulgare"
 /cultivar="CI16155 (Mia13)"
 /db_xref="taxon:4513"
 /clone="HV_C2A006M10f"
 /clone_lib="Hordeum vulgare seedling green leaf EST
 library HVC2A006M10f (Erysiphe infected & control)"
 /library HVC2A006M10f (Erysiphe infected & control)"
 /tissue_type="seedling green leaf"
 /lab_host="TJC121"
 /note="Vector: lambdaDazAP; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 133 a 199 c 173 g 87 t 3 others
 ORIGIN

Query Match 15.9%; Score 187; DB 146; Length 595;
 Best Local Similarity 76.6%; Pred. No. 2e-31;
 Matches 229; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 159 gcggcggcggtatccgcgagcaggaccgctgatgccgatcgcaacgtgatccgcattc 218
 Db 115 GAGCTGTCTGCGCGGGGAGCAGGACCGCTTCTGCCATCGCCACGTCAGCCCGATC 174
 QY 219 atgcgcggcggtgctgcggcgccacgccaagtactctcgacgacctgaaggagacgtcccg 278
 Db 175 ATGAGAGAGCGCTCCCGGCCAACGCCAAGATCATGAAAGACGCCAAGGAGACGGTGCAG 234
 QY 279 gagtcgggtgcgagtaatacatagctttatcacgggggagggaccaacgagcgttgccagcgg 338
 Db 235 GAGTGCCTCTCCGAGTTATCTCTTCATCATCCGCGGAGGCTCCGACAAGTCCACGCCG 294
 QY 339 gacgagcgaagaccatccacgcgagcagctgctgtgggcatagacgccccctgccttc 398

OM of: US-09-435-054-2 to: EST:* out_format : pfs

Date: Oct 25, 2001 2:03 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p2n.model -DEV=xlh
-Q/cgnt2_1/USPTO.spool/US09435054/runat_25102001_085522.6964/app_query.fasta_1.338
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -GAPOP=12.000 -CAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY
-WAIT_THREADS=1

Search information block:

Query: US-09-435-054-2
Query length: 278
Database: EST.*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 1006.350000

score_list:

Sequence	Strd Orig	Zscore	Escore	Len	Documentation
gb_est30:AU088581	+	771.00	1053.30	695	1 AU088581 AU088581 Rice callus C
gb_est30:BE603222	+	598.50	819.72	528	1 BE603222 HVSMH0102116f Hordeum
gb_est51:AW756413	+	481.00	657.74	591	1 AW756413 s12a12.y1 Gm-cl036 G
gb_est102:BG551755	+	465.50	638.18	2.0e-26	1 BG551755 s24a21.y1 Gm-cl075 G
gb_est47:AW432980	+	453.50	637.19	2.3e-26	1 AW432980 si03a01.y1 Gm-cl029 G
gb_est90:BF593304	+	465.50	622.29	1.5e-25	1 BF593304 su76f03.y1 Gm-cl055 G
gb_est28:AL506199	+	446.50	610.45	7.0e-25	1 AL506199 AL506199 Hordeum vulg
gb_est51:AW754604	+	431.00	590.23	9.3e-24	1 AW754604 PC04B12 Pine Triplex F
gb_est54:AW981720	+	422.00	578.07	4.4e-23	1 AW981720 PC15H07 Pine Triplex F
gb_est38:AL509098	+	417.50	573.60	7.9e-23	1 AL509098 AL509098 Hordeum vulg
gb_est74:BE418716	+	410.00	556.23	7.3e-22	1 BE418716 SC1074.B01R990724 ITFC
gb_est27:AL1993140	+	406.50	555.99	7.5e-22	1 AL1993140 701502287 A1. thaliana
gb_est85:BF263449	+	405.00	553.61	1.0e-21	1 BF263449 HV_CEA0006M10f Hordeum
gb_est95:BF263455	+	403.00	553.61	1.0e-21	1 BF263455 HV_CEA0006M16f Hordeum
gb_est39:AV550943	+	403.00	551.43	1.3e-21	1 AV550943 AV550943 Arabidopsis t
gb_est38:AV424305	+	401.00	551.52	1.3e-21	1 AV424305 AV424305 Lotus japonic
gb_est79:BE803572	+	400.00	549.05	1.8e-21	1 BE803572 sr60e11.y1 Gm-cl052 G
gb_est100:BG440251	+	400.00	545.40	2.9e-21	1 BG440251 GA_Ea0006K20f Gossyp
em_estp15:AW666992	+	400.00	545.29	3.0e-21	1 AW666992 GA_Ea0006K20 Gossyp
gb_est69:BE021941	+	397.00	543.80	3.6e-21	1 BE021941 sm64d05.y1 Gm-cl028 G
gb_est24:AL1725612	+	394.00	537.48	8.1e-21	1 AL1725612 BNLGH112445 Six-day CC
gb_est99:BG363233	+	393.50	538.76	6.9e-21	1 BG363233 sac11h11.y1 Gm-cl040 G
gb_est46:AW395227	+	393.50	538.02	7.5e-21	1 AW395227 sh45e04.y1 Gm-cl017 G
gb_est91:BF715909	+	393.00	540.27	5.6e-21	1 BF715909 saalle08.y1 Gm-cl058 G
gb_est50:AW719547	+	392.00	537.71	7.8e-21	1 AW719547 LjNESt6a3r Lotus japon
gb_est50:AW20671	+	392.00	536.90	8.7e-21	1 AW20671 LjNESt6a3r Lotus japon
gb_est49:AW597630	+	392.00	536.30	9.4e-21	1 AW597630 sj96g06.y1 Gm-cl023 G
gb_est47:AW459387	+	391.50	537.49	8.1e-21	1 AW459387 sh23f03.y1 Gm-cl016 G
gb_est100:BG445358	+	391.50	531.49	1.7e-20	1 BG445358 GA_Ea00027N18f Gossyp
em_estp15:AW730639	+	391.50	531.40	1.8e-20	1 AW730639 GA_Ea00027N18f Gossyp
gb_est99:BG314203	+	390.50	533.08	1.4e-20	1 BG314203 WHE2460_E10_12025 Trif
gb_est74:BE441135	+	390.00	532.92	1.4e-20	1 BE441135 EST408405 tomato devel
gb_est71:BE210041	+	388.50	533.61	1.3e-20	1 BE210041 so38b01.y1 Gm-cl039 G
gb_est90:BF597252	+	388.50	530.68	1.9e-20	1 BF597252 su96c06.y1 Gm-cl056 G
gb_est75:BE156510	+	387.50	529.68	2.2e-20	1 BE156510 WHE611_D10_H192a Wheat
gb_est51:AW775623	+	387.00	530.40	2.0e-20	1 AW775623 EST334688 DSTL Medicag
gb_est74:BE413647	+	387.00	527.13	3.0e-20	1 BE413647 SCU001_E10_R990714 ITFC
gb_est51:AW733618	+	386.50	529.17	2.3e-20	1 AW733618 sk75h06.y1 Gm-cl016 G
gb_est21:AL1495007	+	385.50	529.61	2.2e-20	1 AL1495007 sa89f03.y1 Gm-cl004 G
gb_est69:BE060015	+	384.00	529.57	2.2e-20	1 BE060015 sn39h06.y1 Gm-cl027 G

gb_est26:AI900024 + 384.00 524.26 4.4e-20 633 ! AI900024 sb97g11.y1 Gm-cl012
gb_est54:AW980494 + 384.00 523.54 4.8e-20 682 ! AW980494 EST391647 GVN Medic
gb_est51:AW738727 + 383.50 523.91 4.6e-20 611 ! AW738727 EST340154 tomato fl
gb_est51:AW760103 + 383.00 526.78 3.2e-20 422 ! AW760103 sl58b03.y1 Gm-cl027
gb_est49:AW621652 + 382.50 523.89 4.6e-20 531 ! AW621652 EST312450 tomato ro

seq_name: gb_est30:AU088581

seq documentation_block:
LOCUS AU088581 695 bp mRNA EST 31-MAR-2000
DEFINITION AU088581 Rice callus Oryza sativa subsp. japonica cDNA clone C52742
, mRNA sequence.

ACCESSION AU088581
VERSION AU088581.1 GI:7378310
KEYWORDS EST.
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
1 (bases 1 to 695)
Sasaki,T. and Yamamoto,K.
Rice cDNA from callus (2000)
Unpublished (2000)

CONTACT: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT -"RGP"
C52742_12A.

FEATURES
source
1..695
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub_species Japonica"
/db_xref="taxon:39947"
/clone="C52742"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 124 a 245 c 232 g 88 t 6 others
ORIGIN

alignment_scores:
Quality: 771.00 Length: 239
Ratio: 4.016 Gaps: 10
Percent Similarity: 80.335 Percent Identity: 70.293

alignment_block:
US-09-435-054-2 x AU088581 from: 1 to: 695

Align seg 1/1 to: AU088581 from: 1 to: 695
33 ProAlaileArgGluGlnAspArgLeuMetProileAlaAsnValilear 49
|||||
2 CCGGCTATACGTGAGCAGACCGCGGTGATCGGATCGGACGTATCGG 51
49 gileMetArgArgValLeuProAlaHisAlaLysleSerAspAlaL 66
|||||
52 CAPCATCGCGCGGTCTCTCCGCGCAGCCAGCATCTCGGACGCGCA 101
66 yslGluThrileGlnGluCysValSerGluTyrileSerPheilerhrGly 82
|||||
102 AGGAGACGATCAGGAGTCCGTGCGGAGTACATCATCATCACCGG 151
83 GluAlaAsnGluArgCysGlnArgGluGlnArgLysThrileThrAlaGl 99

152 GAGGCCACGACGGTGCACGGCGGACGCGAGACCATCACCGCGCA 201
99 uAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGluP 116
202 GGACGTGCTCTGGGCCATGACGCCCTCGGCTTCGACGACTACGTGCGAC 251
116 roLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg 132
252 CCCTCGGGCTTACTCTCACCGCTACCGCGAGTTCAGGGGGAGATCCCCG 301
133 GlyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisH 149
302 GCGCTCGCGCTCGGCGTGGCGCGCGG.CGCGGCGACCAACCA 345
149 sProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProValS 166
346 TGGTCAC.GTGCGTGGGATGCTCAAGTCCCGCGCGACGGCT 386
166 erGlyAlaAlaMetLeuProHisHisHisHisAspMetGlnMet 182
387 CCATGGTGACG.CACCACGACATGCAGATG 415
183 HisAlaAlaMetTyrGlyGlyThrAlaValProProProAlaGlyProPr 199
416 CACGGCGGCATGTACGGTGGCGCGCGGTGCCGCGNCGCGCATCTCC 465
199 o. HisHisGlyGlyPhe.LeuMetProHisProGlnGlySerSe 213
466 TCGNCCACCGCGCTTCCACCACTCATGCG.GC 500
213 rHisTyrLeuProTyrAla.TyrGluProThrTyrGlyGlyGluH 228
501 GCACACCGCCCGAGTACGGCGCGCGCGTACGAC.ATGTACGGCGGCGAG 547
228 iSalMetAlaAlaTyrTyrGlyGlyAlaAlaTyrAlaProGlyAsnGly 244
548 ACGGGATGGCGGTACTACGGCGG.ATGTACGGCGCGCGCGACGCG 594
245 GlySerGlyAspGlySerGlySerGlyGlyGlyGly.Se 258
595 GGCNACGGGCGGCGAGCAGCGGCGGTGGCGCGCGGACCGCGGNAGAA 644
258 rAlaSerHisThrPro 263
645 CGTCAATTNACACCA 660

seq_documentation_block:					
LOCUS	BE603222	528 bp	mRNA	EST	02-MAR-2001
DEFINITION	HVSMH0102J16f	Hordeum vulgare 5-45 DAP spike EST library			
	HVCNNA0009	(5 to 45 DAP)	Hordeum vulgare cDNA clone HVSMER0102J16f, mRNA sequence.		
ACCESSION	BE603222				
VERSION	BE603222.2	GI:13191083			
KEYWORDS	EST.				
SOURCE	barley.				
ORGANISM	Hordeum vulgare				
	Eukaryota; Viridiplantae;	Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta;	Liliopsida; Poales; Poaceae; Pooideae			
	; Triticeae; Hordeum.				
REFERENCE	1	(bases 1 to 528)			
AUTHORS	Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and Wood,T.				
TITLE	Development of a genetically and physically anchored EST resource for barley genomics				
JOURNAL	Unpublished (2000)				
COMMENT	On Aug 21, 2000 this sequence version replaced gi:9860783. Contact: Wing RA Clemson University Genomics Institute				

```

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7286
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: AATTAACCTCACTAAAGGG
High quality sequence stop: 512.

FEATURES
    source
        1..528
            /organism="Hordeum vulgare"
            /cultivar="Morex"
            /db_xref="taxon:4513"
            /clone="HVSMEH0102J16f"
            /clone_lib="Hordeum vulgare 5-45 DAP spike EST library"
            /vcdna="HVCNA0009 (5 to 45 DAP)"
            /tissue_type="5-45 DAP Spike"
            /lab_host="SOLR"
            /note="vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT      97 a 181 c 182 g 68 t
ORIGIN

alignment_scores
    Quality: 598.50
    Ratio: 4.215
    Percent Similarity: 81.609
    Gaps: 5
    Percent Identity: 71.264

alignment_block:
    US-09-435-054-2 x BE603222

Align seg 1/1 to: BE603222 from: 1 to: 528

```

361 GGAAGACCATACCCGAGGACGCTGCTTTGGCCATCAGCAGCTTGA 410

110 PheAspAspTyrValcIuProLeuGlyAlaTyrLeuHisArgTyrArgG1 126

411 TTCGACGACTACATGACACCGTTGACCATGTACCTTCACCGCTACCGTGA 460

126 uPheGluGlyAsp 130

461 ACTTGAGGCTGAC 473

seq_name: gb_est47:AW432980

seq_documentation_block:

LOCUS AW432980 551 bp mRNA 18-JUL-2000

DEFINITION si03a01.y1 Gm-cl029 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl029-97 5' similar to TR:081130 O81130 CCAAT-BOX BINDING FACTOR HAP3 HOMOLOG. ; mRNA sequence.

ACCESSION AW432980

VERSION AW432980.1 GI:6964287

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine. 1 (bases 1 to 551)

REFERENCE

AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpeliding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R/Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com Insert Length: 1007 Std Error: 0.00 High quality sequence stop: 455.

FEATURES

Location/Qualifiers

1..551

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl029-97"

/clone_lib="Gm-cl029"

/tissue_type="very young cotyledons of greenhouse grown plants"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from very young cotyledons (20-50mg fresh weight) of greenhouse grown plants. The library was prepared using the Life Technologies pSuperScript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly(dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

156 a 147 c 144 g 104 t

BASE COUNT

ORIGIN

alignment_scores:

Quality: 465.50 Length: 121

Ratio: 4.392 Gaps: 1

Percent Similarity: 87.603 Percent Identity: 74.380

alignment_block:

US-09-435-054-2 x BG551755 ..

Align seg 1/1 to: BG551755 from: 1 to: 497

10 AlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGlyAl 26

138 GCATCATCATGATCAGTCAGTCACGACGACGAGGAGGACGAA..... 179

26 aAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArgLeuMetp 43

180TGCACGGTCAGGAGGAGCAAGCAGGTTTCATGC 210

43 rolleAlaAsnValIleArgIleMetArgqValIleProAlaHisAla 59

211 CAATCGCTAACGTGATTAGGATCATGCGCAAGATTCTCCCTCCACACCA 260

60 LysIleSerAspAlaLysGluThrIleGlnGluCysValSerGluTy 76

261 ARAATCTCGAGCATGCAAAAGCAACATCCAAAGAGTCGTCTGTAGTA 310

76 rIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluGlnA 93

311 CATCATGCTTCATCACAGGTGAGCGCAACGAGCGTTGCCAGAGGGAGCAG 360

93 rGlySerThrIleThrAlaGluAspValLeuTrrpAlaMetSerArgLeuGly 109

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com High quality sequence stop: 370.

Location/Qualifiers

1..497

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl075-669"

/clone_lib="Gm-cl075"

/tissue_type="differentiating somatic embryos cultured on MSM6AC"

/lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2: XhoI; The cDNA library was constructed from mRNA isolated from differentiating somatic embryos cultured on MSM6AC. The library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into E.coli Electromax DH10B host cells. Tissue culture and library construction were performed by Francoise Thibaud-Nissen and Anu Khana (Lila Vodkin lab, University of Illinois)."

146 a 130 c 131 g 90 t

BASE COUNT

ORIGIN

448 ACTTGAG 454

seq_name: gb_est28.AL506199

seq_documentation_block:

LOCUS AL506199 594 bp mRNA EST 04-JAN-2001
 DEFINITION Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
 Hordeum vulgare cDNA clone HY02F18T 5', mRNA sequence.

ACCESSION AL506199

VERSION AL506199.1 GI:12032414

KEYWORDS

SOURCE

ORGANISM

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

; Triticeae; Hordeum.

1 (bases 1 to 594)

Michalek, W., Weschke, W., Pleissner, K.-P. and Graner, A.

EST sequencing and analysis in barley

Unpublished (2000)

JOURNAL

COMMENT

Contact: Michalek W

Institute for Plant Genetics and Crop Plant Research

Corrensstr.3, D-06466 Gatersleben, Germany

Email: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de

Seq primer: T3 primer for 5' end.

FEATURES

SOURCE

1..594

/organism="Hordeum vulgare"

/cultivar="Barke"

/db_xref="taxon:4513"

/clone_lib="Hordeum vulgare Barke developing caryopsis

(3.-15.DAP)"

/tissue_type="developing caryopsis (3.-15.DAP)"

/lab_host="XLOUR"

/note="Vector: plasmid pBK-CMV; Site.1: EcoRI; Site.2:

XhoI; mRNA was made from developing caryopsis (3.-15.DAP)

of spring barley variety 'Barke', a high quality malting

variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI

(3'-end of cDNA). NOTE: Due to a cloning artefact caused

by the kit, in most cases the EcoRI site is NOT present,

as well as the EcoRI adapter. Average insert size is 1 kb

Sequence trimming: Vector sequences and sequence ends were

trimmed from the 5'- and 3'-end until a 50 bp window

contains less than two ambiguities. The maximum length was

set to 700 bp"

BASE COUNT 109 a 213 c 185 g 85 t 2 others

ORIGIN

alignment_scores:

Quality: 446.50 Length: 193

Ratio: 3.212 Gaps: 4

Percent Similarity: 72.021 Percent Identity: 53.368

alignment_block:

US-09-435-054-2 x AL506199 ..

Align seg 1/1 to: AL506199 from: 1 to: 594

9 AlaAlaGlyAlaGlnGlySerAlaAlaGlyAlaAsnGlyG1 25

58 GCCCTAGCCATCGACGACTTCAACGTCGCCGAGCAGCA.....CGCGC 101

25 yAlaAlaGlnGlnHisAlaAlaProAlaAlaArgGluGlnAspArgLeuM 42

102 GCCGGGCCCACTCAGCGCAGCCGGTGTGGGAGCAGCAGCGCTGA 151

42 etProTleAlaAsnValIleArgIleMetArgValLeuProAlaHis 58

152 TGCCGATCGCAACGATGATCCGCATCATCGCGTGCCTCCCTGCCAC 201

59 AlaLyIleSerAspAspAlaLysGluThrIleGlnGluCysValSerG1 75

|||||
 202 GCCAAGATCTCGACGACGCCAAGGAGCGATCCAGGAGTGGTGTCCGA 251

75 wTyTleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluG 92

252 GTTCATCAGCTTCGTACCGGAGGCAACAGGCGGTGCCACATGGAGC 301

92 lnArgLySThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeu 108

302 ACCGCAAGACCTCAACGCGGAAGACATCGTGTGGGCGCTGAACCGCCTC 351

109 GlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrAr 125

352 GCGTTCGAGACTAGCTCTGCGCCCTCAGCGCTTCCTGCACCGCATGGC 401

125 gGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAlaAlaP 142

402 CAGAGCGGAGGGGGGACAGGTGGTGGCGGTGCAGCGCAACACCGCGCG 451

142 roSerArgGlyGlyAspHisHisProHisSerMetSerPro..... 155

452 CCACGAGGGCGCTCCCGCGCGCGGCC.GNCCGCTCTCCACGCGCTGCC 500

156AlaAlaMetLeuLysSerArgGlyProValSerGlyAlaAl 169

501 GTTGCAGGTACTGCAAGCGTCCATGTACGCGCCCGCTCTCCGTTGCAGC 550

169 aMetLeuProHisHisHisHisAspMetMetHisMetHisAlaAlaM 186

551 T.....GAGCAATCAGATTGCNA 567

186 etTyrGlyGlyThrAlaValProProPro 195

568 GA.....GGCCAGTTGTACGCGCGCCCC 590

seq_name: gb_est51:AW754604

seq_documentation_block:

LOCUS AW754604 535 bp mRNA EST 01-MAY-2000

DEFINITION PC04B12 Pine Triplex pollen cone library Pinus taeda cDNA clone

PC04B12, mRNA sequence.

ACCESSION AW754604

VERSION AW754604.1 GI:7676324

KEYWORDS EST.

SOURCE loblolly pine.

ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.

Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.

The Pine Gene Discovery Project

Unpublished (1999)

Contact: Ross Whetten

Forest Biotechnology Group

North Carolina State University

Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh

NC, 27695-8008

Tel: 919-515-7800

Fax: 919-515-7801

Email: rosswhetten@unity.ncsu.edu

Seq primer: 5' lambda Triplex2 Sequencing Primer.

Location/Qualifiers

1..535

/organism="Pinus taeda"

/db_xref="taxon:3352"

/clone_lib="Pine Triplex pollen cone library"

/dev_stage="immature"

/lab_host="E. coli BM25.8"

/note="Organ: pollen cone; Vector: Lambda Triplex; Site.1:

SfiI (A); Site.2: SfiI (B); Immature pollen cones were

collected in the early spring, frozen and used for mRNA

isolation. The SMART-PCR method (Clontech) was used to

Source

Prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 143 a 103 c 129 g 139 t 21 others
ORIGIN

alignment_scores:
Quality: 431.00 Length: 99
Ratio: 4.585 Gaps: 0
Percent Similarity: 94.949 Percent Identity: 82.828

alignment_block:
US-09-435-054-2 x AW754604 ..

Align seg 1/1 to: AW754604 from: 1 to: 535

```

34 AlAlieArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgII 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 GCTGTGAGAGCAAGATAGTGTATGCCCCATTCCTAATGTCATTAGGAT 237
50 eMetArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysG 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 AATGAGGAAGTCTTACCCACCCATGCAAGATTTCTGATGATGCAAGG 287
67 luThrIleGlnCysValSerGluTyrIleSerPheIleThrGlyGlu 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
288 AGACCATTCGAAGATGTGTCTGATGATACATAAGCTTCATCACCAGTGA 337
84 AlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAs 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 GCGAAGCATGCTGCGCAAGAGGAGCAGAGAAAACACTATCATCTGCTGA 387
100 pValLeuTrpAlaMetSerArgLeuGlyPheAspTyrValGluProL 117
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 TGCTCTCGGCTATGACCAAGTTGGGTTTGTGATGATTATGTGGAGCAT 437
117 euGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg 132
|||:|||||:|||||:|||||:|||||:|||||:|||||:
438 TACATTTTACCTCCAGAAATACAGAGATGCTGGAAGNNNANNNCAGA 484

```

seq_name: gb_est54:AW981720

seq_documentation_block: 526 bp mRNA EST 02-JUN-2000
LOCUS AW981720
DEFINITION PC15H07 Pine triplex pollen cone library Pinus taeda cDNA clone
PC15H07, mRNA sequence.
ACCESSION AW981720
VERSION AW981720.1 GI:8173288
KEYWORDS EST.
SOURCE loblolly pine.
ORGANISM Pinus taeda

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferales; Pinaceae; Pinus; Pinus.
1 (bases 1 to 526)
Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
The Pine Gene Discovery Project
Unpublished (1999)
Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
, NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu

Seq primer: 5' lambda Triplex2 Sequencing Primer.
source location/Qualifiers
I. .526
/organism="Pinus taeda"
/db_xref="taxon:3352"
/clone="PC15H07"

FEATURES
source

/clone_lib="Pine Triplex pollen cone library"
/dev_stage="immature"
/lab_host="E. coli BM25.8"
/note="Organ: pollen cone; Vector: Lambda Triplex; Site: 1:
Sfil (A); Site: 2: Sfil (B); Immature pollen cones were
collected in the early spring, frozen and used for mRNA
isolation. The SMART-PCR method (Clontech) was used to
prepare a library from 1 ug total RNA, using the Lambda
Triplex vector. Plasmid subclones in pTriplex were
recovered by cre-lox excision in E. coli strain BM25.8 and
sequenced from the 5' end."

BASE COUNT 176 a 100 c 122 g 128 t
ORIGIN

alignment_scores:
Quality: 422.00 Length: 121
Ratio: 3.944 Gaps: 0
Percent Similarity: 88.430 Percent Identity: 68.595

alignment_block:

US-09-435-054-2 x AW981720 ..

Align seg 1/1 to: AW981720 from: 1 to: 526

```

9 AlaAlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGI 25
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 GCTTCAATGGCAGCAAGATGCAAGCCCAACCCAGCAGATACAGACCTC 211
25 yAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArgLeuM 42
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
212 CGAGGATGGAGATAGAGAAAATTTGCTGTAGAGAAACAGACAGGTTTA 261
42 etProIleAlaAsnValIleArgIleMetArgArgValLeuProAlaHis 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
262 TGCTATCGCAATGTATCAGGATCATGAGAAAGTCTCCCAACACAT 311
59 AlAlaLysSerAspAlaLysGluThrIleGlnGluCysValSerGI 75
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 GCGAAAATTTCTGATGATGCAAGAAACAAATACAGGAATGTGTTTCAGA 361
75 uTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgGluG 92
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 ATTTATCAGTTTCATCAGTGTAGAGCAATGACCGTGCCAAAAGGAAC 411
92 lnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArgLeu 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
412 AAAGGAAGACTATCCTGCGGAGGATGTTTATGCGCCATGAATAAGTTA 461
109 GlyPheAspTyrValGluProLeuGlyAlaTyrLeuHisArgTyrAr 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
462 GGTTTTGACAATTACATGGATCCATTCAGTCTTACCTTCAGAGATATAG 511
125 gGluPheGluGly 129
|:|||||:
512 AGGTATGAGGGT 524

```

seq_name: gb_est28:AL509098

seq_documentation_block: 441 bp mRNA EST 04-JAN-2001
LOCUS AL509098
DEFINITION Hordeum vulgare Barke developing caryopsis (3.-15.DAP)
Hordeum vulgare cDNA clone HV10L07V 5', mRNA sequence.
ACCESSION AL509098
VERSION AL509098.1 GI:12035601
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
1 (bases 1 to 441)
Michalek, W., Weschke, W., Pleissner, K.-P. and Graner, A.

[illegible]

seq_name: qb_est27:AI995140

seq_documentation_block:		575 bp	EST	08-SEP-1999
LOCUS	AI995140		mRNA	
DEFINITION	701502287 A. thaliana, Ohio State clone set Arabidopsis thaliana cDNA clone 701502287, mRNA sequence.			

ACCESSION	AI995140
VERSION	AI995140.1
KEYWORDS	EST.
SOURCE	thale cress.
ORGANISM	<i>Arabidopsis thaliana</i>

Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Streptophyta; Magnoliophyta; eudicotyledons; core eudicots; Spermatophyta; Magnoliopsida; eurossids II; Brassicales; Brassicaceae; Arabidopsis. Rosidae; eurossids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 575)
 REFERENCES
 AUTHORS
 Chen, J., Momiyama, M., Chan, E., Mooney, M., Carroon, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Brzoska, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Borillo, C., Carpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Nobrega, A., Murrey, L., Turner, C., Krlorian, S., Elder, B. and Hanson, D.
 Arabidopsis thaliana Gene Expression MicroArray
 TITLE

JOURNAL COMMENT
Unpublished (1999)
Contact: David Smoller, ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte

Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com.

[illegible]

```

1. :373
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/cclone="701502087"

```

```
/clone="70150228/"
/clone_lib="A. thaliana, Ohio State clone set"
/note="CDNA library was made from selected clones from the
Arabidopsis thaliana Ohio State clone set."
```

	BASE COUNT						ORIGIN					
	154 a		116 c		177 g		118 t		state crone s		10 others	
			<i>Atariopsis</i>		<i>Chiliana</i>		<i>Olio</i>					

alignment_scores:		
Quality:	406.50	Length: 186
Ratio:	2.863	Gaps: 4
Percent Similarity:	76.344	Percent Identity: 46.237

alignment_block: US-09-435-054-2 x AT995140

Align seq 1/1 to: AI995140 from: 1 to: 575

```

18 AlaGlyGlyAlaAsnAsnGlyGlyAlaAlaGlnGlnHisAlaAlaProAl 34
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
11 TCCGGTTGGAGGGCAAAACGGGAACAACACGAGCGACATCCTCTCTGTC 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 aileArgGluAlaSpArgLeuMetProIleAlaAsnValIleArgIleM 51
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 TCCAAAGAGAGCAGACAGGTCTTGTGGCATGCGTCAAGTCAGCCCGGATCA 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
51 etArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGlu 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
111 TGAAGAAGGCCCTGCCCGCCCAACGCCAAGATCTCTAAACATGCCAAAGAG 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
68 ThrIleGlnCysValSerGluTyrIleSerPheIleThrGlyGluAl 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
161 ACGATTGCAGAGGATGTTGTTCCGAGTTCATCAGCTTCGTCACCGGAGAAGC 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
84 aAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAspV 101
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
211 ATCTGATAAGTGTGAGAAGGAGGAAGGAGGATCAACGGAGACGAT 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
101 alLeuTyrPalaMetSerArgLeuGlyPheAspAspTyrValGluProLeu 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
261 TGCCTCTGGCTATGACTACTCTAGGTTTTCAGGATTATCTTCAGCCATTG 310
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 GlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArgGlyVa 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
311 AAAGTTTACTTGCAGAGGTTTAGCGAGATCGAAGGGAGGAGACTGGACT 360
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
134 lGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHisProH 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
361 AGGAGG.....GCACAGACTGTGTGTGAGGTCGGAGAGC 394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 iSerMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGly 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
395 ATCAGACAGATGCTGTCGA.....GATGGCGGTGGGTCTACAGGT 435
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
168 AlaAlaMet...LeuProHisHisHisHisHisHisAspMetGlnMethi 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
436 NNGTGTNGTGGGATGCAGTATCACCACATCATCAGTTTCTTTCACCCAGCA 485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 sAlaAlaMetTyrGlyGlyThrAla.....V 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
486 GACCCATATGATTGTAGCCACANNGTGGCGTACGACAGTNGANGTNGAGC 535
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
192 alProPro 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

536 TGCTCCG 543
seq_name: gb_est85:BF263449
seq_documentation_block:
LOCUS   BF263449      595 bp      mRNA      EST      09-MAR-2001
DEFINITION   HV_CEA0006M10f Hordeum vulgare seedling green leaf EST library
HVCNDA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
HV_CEA0006M10f, mRNA sequence.
ACCESSION   BF263449
VERSION     BF263449.2  GI:13260832
KEYWORDS    EST.
SOURCE      barley.
ORGANISM    Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 595)
AUTHORS    Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
TITLE       Development of a genetically and physically anchored EST resource
            for barley genomics
JOURNAL     Unpublished (2000)
COMMENT     On Nov 17, 2000 this sequence version replaced gi:11194443.
            Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACTCTCACTAAAGGG
            High quality sequence stop: 594.
            Location/Qualifiers
FEATURES             source
    seq_name: gb_est85:BF263449
    seq_documentation_block:
    LOCUS   BF263455      595 bp      mRNA      EST      09-MAR-2001
    DEFINITION   HV_CEA0006M16f Hordeum vulgare seedling green leaf EST library
    HVCNDA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
    HV_CEA0006M16f, mRNA sequence.
    ACCESSION   BF263455
    VERSION     BF263455.2  GI:13260837
    KEYWORDS    EST.
    SOURCE      barley.
    ORGANISM    Hordeum vulgare
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
    ; Triticeae; Hordeum.
    REFERENCE   1 (bases 1 to 595)
    AUTHORS    Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
    TITLE       Development of a genetically and physically anchored EST resource
            for barley genomics
    JOURNAL     Unpublished (2000)
    COMMENT     On Nov 17, 2000 this sequence version replaced gi:11194449.
            Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACTCTCACTAAAGGG
            High quality sequence stop: 590.
            Location/Qualifiers
FEATURES             source
    14 AsnGlySerAlaAlaGlyGlyAlaAsnAsnGlyGlyAlaAlaGlnGlnHi 30
    79 GAACAAGACTCCAGTGGCGGAGCAACAGCGGGG.....GAGCT 119
    30 sAlaAlaProAlaIleArgGluGlnAspArgLeuMetProIleAlaAsnV 47
    120 CTCGTCGCGG.....CGGAGGAGGAGCGCTTCCTGCCCATCGCCCAAGC 163
    47 allleArgIleMetArgValLeuProAlaAlaHisAlaLysIleSerAsp 63
    164 TCAGCGCGGATCATGAAGAGCGCTCCGCGCAACGCCAAGATCAGCAAA 213
    64 AspAlaLysGluThrIleGlnGluCysValSerGluTyrIleSerPheI 80

```



```

126 uPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAlaAlaProS 143
|:::|||||:::||||| ||| |||:::|
480 GATGAGGGTGACACTAAGGATCAGGA.....AAGGCGGGGAATCGA 523
|
143 erArgGlyGlyAspHisHisProHisSerMetSer 154
|| |||:::|||||:::|||||
524 CTGCAAGAGAGATGGTCAACCAAGCCCAAGTGTCT 558

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 25, 2001, 10:49:35 ; Search time 42.84 Seconds
(without alignments)
5183.519 Million cell updates/sec

Title: US-09-435-054-1
Perfect score: 1173
Sequence: 1 ccacgcgtccgcaccacac.....tcaaaaaaaaaaaaaaaaaa 1173

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 9465562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	179	15.3	3395	4	US-09-103-478-3
2	179	15.3	7560	4	US-09-103-478-4
3	177	15.1	627	4	US-09-103-478-1
4	78.4	6.7	835	4	US-09-319-989-3
5	65.6	5.6	2493	1	US-07-977-434-5
6	65.6	5.6	2493	1	US-08-458-819-5
7	65.6	5.6	2493	5	PCT-US91-07035-5
8	56	4.8	28958	1	US-08-258-261B-6
9	56	4.8	28958	1	US-08-456-837-6
10	56	4.8	28958	1	US-08-457-342-6
11	56	4.8	28958	1	US-08-457-646A-6
12	56	4.8	28958	1	US-08-458-076A-6
13	56	4.8	28958	1	US-08-764-233A-4
14	56	4.8	28958	1	US-08-457-335A-6
15	56	4.8	28958	3	US-08-729-214-6
16	56	4.8	28958	3	US-09-028-934-6
17	56	4.8	49377	1	US-08-764-233A-1
18	55.4	4.7	5970	6	US-09-320-878-21
19	55.2	4.7	4897	6	5196516-7
20	55.2	4.7	20235	3	US-07-642-734C-3
21	55.2	4.7	20235	3	US-08-439-009A-3
22	54.6	4.7	4257	2	US-08-890-473-1
23	54.6	4.7	4257	4	US-09-259-821A-1
24	54.6	4.7	4257	4	US-08-843-659-1
25	54.6	4.7	12001	1	US-08-458-568A-11
26	54.2	4.6	38506	3	US-09-320-878-19
27	53.2	4.5	1965	4	US-09-178-252-26

28	52.8	4.5	4826	4	US-09-192-983-3
29	52.2	4.5	2539	3	US-09-000-016-3
30	52.2	4.5	2809	3	US-09-000-016-1
31	51.6	4.4	675	4	US-08-998-416-723
32	51.6	4.4	30001	1	US-08-125-468-1
33	51.6	4.4	30001	2	US-08-474-933-1
34	51.4	4.4	11219	1	US-07-642-734C-1
35	51.4	4.4	11219	3	US-08-439-009A-1
36	51	4.3	1212	6	5212296-17
37	51	4.3	1998	6	5212296-8
38	50.6	4.3	833	2	US-08-403-852D-3
39	50.6	4.3	833	3	US-08-510-648B-3
40	50.6	4.3	833	4	US-09-231-818-3
41	50.6	4.3	5392	2	US-08-403-852D-1
42	50.6	4.3	5392	3	US-08-510-646B-1
43	50.6	4.3	5392	4	US-09-231-818-1
44	50.4	4.3	2244	1	US-08-476-519-10
45	50.4	4.3	2244	5	PCT-US95-09323-10

ALIGNMENTS

RESULT 1
US-09-103-478-3
Sequence 3, Application US/09103478
Patent No. 6235975
GENERAL INFORMATION:
APPLICANT: Harada, John
APPLICANT: Lotan, Tamar
APPLICANT: Ohto, Masa-aki
APPLICANT: Goldberg, Robert B.
APPLICANT: Fischer, Robert L.
TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09103,478
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/026,221
FILING DATE: 19-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/804,534
FILING DATE: 21-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Einhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-077611US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3395 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-103-478-3

Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 723, App
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Patent No. 5212296
Patent No. 5212296
Sequence 3, Appli
Sequence 3, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 1, Appli
Sequence 10, Appli
Sequence 10, Appli


```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/026,221
: FILING DATE: 19-FEB-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/804,534
: FILING DATE: 21-FEB-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Einhorn, Gregory P.
: REGISTRATION NUMBER: 38,440
: REFERENCE/DOCKET NUMBER: 023070-07761110S
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 627 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..627
: OTHER INFORMATION: /product= "LECI"
US-09-103-478-1

```

	Query Match	15.1%;	Score 177;	DB 4;	Length 627;
	Best Local Similarity	68.1%;	Pred. No. 2.2e-27;		
	Matches 246;	Conservative 0;	Mismatches 115;	Indels 0;	Gaps 0;
QY	197	ccggcgcgagagaatggtctcggcgcgcgggcgccaacatggcgcgctgtctcagcagc	156		
Db					
Db	5	CGAGCTCAGTATAGTACGGCCGCGGTGCACAAGAACAATGGTATCGTGTCGACGACG	64		
QY	157	atgcggcgccggcggtatccgcgagcagagcggctgatgcccatctcgaaacgtgaaccga	216		
Db					
Db	65	AACCACCATTGTGTGCTGTGAGCAAGAACCAATACATGCCAATCGGAACGTCATAAGAA	124		
QY	217	tcatgcgcgcgcgtgcgcgcgcgcacgcccaagatctcggacgcgcgaaggagacgattcc	276		
Db					
Db	125	TGATCGGTAAAACCTTACCGTGCTCACGCCAAAAATCTCTGACGACGCGCAAAGAAACGATT	184		
QY	277	aggaatgcgcgtcggagttacatcagcttcatacgggggaagcccaacgagcggtgcccagc	336		
Db					
Db	185	ANGATGTGTTCCGAGTTACATCAGCTTCGTGACCGGTGAAGCCAACGACGCTTGCCAAC	244		
QY	337	gggagcagcgaagaccataccccgcggagacgtgctgtggccatgagccgctcgctt	396		
Db					
Db	245	GTGAGCAACGTAAGACCATAACTGCTCAAGATATCCTTTGGGCTATGACCAAGCTTGGGT	304		
QY	397	tgcagcactacgtcagccgctcgcgccttaacctcaacgcctaccgcgagttcgaggcg	456		
Db					
Db	305	TCGATAACTAGTGAGCCCCCTCACCGTGTTCAATTAAACCGTTACCGTGAGTAGAGACCG	364		
QY	457	a	457		
Db					
Db	365	A	365		

RESULT⁴
US-09-319-989-3
; Sequence 3, Application US/09319989
; Patent No. 6190914
; GENERAL INFORMATION:
; APPLICANT: Grivell, Leslie A.
; APPLICANT: Teixeira De Mattos, Maarten J.
; APPLICANT: Blom, Jolanda
; TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
; TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
; TITLE OF INVENTION: METHODS
; FILE REFERENCE: 24615-20123.00
; CURRENT APPLICATION NUMBER: US/09/319,989

```

; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/NL97/00688
; EARLIER FILING DATE: 1997-12-12
; EARLIER APPLICATION NUMBER: EPO 96203520
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA construct
; OTHER INFORMATION: encoding HAP3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(632)
US-09-319-989-3

```

[illegible]

RESULT 5
US-07-977-434-5
; Sequence 5, Application US/07977434
; Patent No. 5456591
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: WordPerfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/977,434
; FILING DATE:

Db 16596 CTGGGGCGACAGGCTCTCCATCGCGCGCTCAACAGCCCCAGGGCCACGCTCGTATCCGG 16655

QY 656 cgggctctctccacacgcggttcttc 683

Db 16656 CGAGCCCGCGCGCTCGACGCGCTGATC 16683

RESULT 9

US-08-456-837-6
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456.837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995

; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-456-837-6

Query Match 4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgccgagacgtgctgtgggacatagcgcctcgcttcgactcagcactcagtcgagcc 415

Db 16356 CGCCCTCTTTCGGGTCATGGTCTCCCGCGCGCTCGTGGCGCTCGCGCTCGAGCC 16415

QY 416 gctcgcgctactctccaccgctaccgaggttcgagggcgacgcgcgcggtcggct 475

Db 16416 CGCGCGCGTCTCGGCGCACAGCCAGCGGAGATAGCGCGCTCTGTCGAGCGCTCT 16475

QY 476 cgtcccgggggcgcccccacgcgcggcgagccaccaccacccacactcactcgcagc 535

Db 16476 CTCCTCGAGGAGCGCGCGCATCGCGCCCTCGCGAGCAAAAGCGTCAACACCGTCGCC 16535

QY 536 ggcgatgtcaagtcccgcgggccagtcctccggagcgccatgtaccgcaccaccacca 595

Db 16536 GGCAACGGGATGCGCGCTCGAGTCTCGCGCTCCGACCTCCAGACCTACCTCGCTCC 16595

QY 596 ccaccacgacatgcagatgcacgcgcgcctacgtacg9gggagcgcgcgtgccccgcgcgc 655

Db 16596 CTGGGGCGACAGGCTCTCCATCGCGCGCTCAACAGCCCCAGGGCCACGCTCGTATCCGG 16655

QY 656 cgggctctctccacacgcggttcttc 683

Db 16656 CGAGCCCGCGCGCTCGACGCGCTGATC 16683

RESULT 10

US-08-457-342-6
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995

; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO

```
; ANTI-SENSE: NO
US-08-457-342-6

Query Match          4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgcgagagctgctgtggccatgagcgcctcgcttgcagactacgtcagacc 415
Db 16356 GCGCCCTTTTGGCGTCATGTCTCCCTGGCGGCCCTCTGCGCTCGCTCGGCTCGAGCC 16415
QY 416 gctcggcgctactccacgcctaccgcaggttcgagggcgagcgcgcggcgctcggtc 475
Db 16416 CGCGCGCGTCTCGGCGCACAGCCAGGCGGAGATAGCGCGCGCTTCGTGCGAGCGCTCT 16475
QY 476 cgtcccgggggcgcccccacgcgcgcggcgagcgcacccacccacactccatgtcgccagc 535
Db 16476 CTCCCTCGAGGAGCGCGCGCATCGCGCCCTTGCAGCAAGAGCGTCCACCGCTCGCC 16535
QY 536 ggcgatctcaagtcccgcgggccagctctccgagcgcgcctaccgcacaccacacca 595
Db 16536 GCGAACGGGATGCGCGCCCTCGAGCTCGCGCTCGGCTCGGACTCCAGACCTACCTCGCTCC 16595
QY 596 ccaccacagacatgcagatgcacgcgcgcgcgtatgcgggggaaacgcgcgtgccccgcgcgc 655
Db 16596 CTGGGCGGACAGGCTCTCCATCGCGCGCTCAACAGCCCCAGGCGCAGCGTCTGTATCCGG 16655
QY 656 cgggcctctcaccacggcggttctctc 583
Db 16656 CGAGCGCGCGCGCTCGACGCGCTGATC 16683

RESULT 11
US-08-457-646A-6
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
```

```
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-457-646A-6

Query Match          4.8%; Score 56; DB 1; Length 28958;
Best Local Similarity 48.2%; Pred. No. 0.0056;
Matches 158; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

QY 356 caccgcgagagctgctgtggccatgagcgcctcgcttgcagactacgtcagacc 415
Db 16356 GCGCCCTTTTGGCGTCATGTCTCCCTGGCGGCCCTCTGCGCTCGCTCGGCTCGAGCC 16415
QY 416 gctcggcgctactccacgcctaccgcaggttcgagggcgagcgcgcggcgctcggtc 475
Db 16416 CGCGCGCGTCTCGGCGCACAGCCAGGCGGAGATAGCGCGCGCTTCGTGCGAGCGCTCT 16475
QY 476 cgtcccgggggcgcccccacgcgcgcggcgagcgcacccacccacactccatgtcgccagc 535
Db 16476 CTCCCTCGAGGAGCGCGCGCATCGCGCCCTTGCAGCAAGAGCGTCCACCGCTCGCC 16535
QY 536 ggcgatctcaagtcccgcgggccagctctccgagcgcgcctaccgcacaccacacca 595
Db 16536 GCGAACGGGATGCGCGCCCTCGAGCTCGCGCTCGGCTCGGACTCCAGACCTACCTCGCTCC 16595
QY 596 ccaccacagacatgcagatgcacgcgcgcgcgtatgcgggggaaacgcgcgtgccccgcgcgc 655
Db 16596 CTGGGCGGACAGGCTCTCCATCGCGCGCTCAACAGCCCCAGGCGCAGCGTCTGTATCCGG 16655
QY 656 cgggcctctcaccacggcggttctctc 583
Db 16656 CGAGCGCGCGCGCTCGACGCGCTGATC 16683

RESULT 12
US-08-458-076A-6
; Sequence 6, Application US/08458076A
; Patent No. 5698425
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; MEDIUM TYPE: Floppy disk
```


Query Match	4.8%	Score 56;	DB 1;	Length 28958;
Best Local Similarity	48.2%;	Pred. No. 0.0056;		
Matches 158;	Conservative	0;	Mismatches 170;	Indels 0; Gaps 0;

QY	356	cacgcgcgagacgtgctgtgggcca	gagccgcctcggcttcgacgactacgtcgagcc	415
Db	16356	CGCCCTCTTTGCCGTATGFTCTCCCTGGCCGCTCTGGCGCTCGCTCGCGCTCGAGCC	16415	
QY	416	gctcggcgctacctcacgcgctaccgcaggttcgaggggcagcgcgcgcgcgtcgggct	475	
Db	16416	CGCCCGCGTCTCGGCGCACACGACAGGGCGGAGATAGCGCCGCCCTTCGTCGCGAGGGGCTCT	16475	
QY	476	cgtcccgggggcgcccccacatcgcgcgcgcgacacacacccgcactccatgtcgccagc	535	

```

: Sequence 6, Application US/08457335A
: Patent NO. 5723759
: GENERAL INFORMATION:
: APPLICANT: Schupp, Thomas M.
: APPLICANT: Ligon, James M.
: APPLICANT: Beck, James Joseph
: APPLICANT: Hill, Dwight Steven
: APPLICANT: Ryals, John Andrew
: APPLICANT: Gaffney, Thomas Deane
: APPLICANT: Lam, Stephen Ting
: APPLICANT: Hammer, Phillip E.
: APPLICANT: Uknes, Scott Joseph
: TITLE OF INVENTION: Genes for the synthesis of
: TITLE OF INVENTION: antipathogenic substances
: NUMBER OF SEQUENCES: 22
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: NY
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/457,335A
: FILING DATE: 01-JUN-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/457,205
: FILING DATE: 01-JUN-1995
: APPLICATION NUMBER: 08/258,261
: FILING DATE: 08-Jun-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Elmer, James Scott
: REGISTRATION NUMBER: 36,129
: REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 919-541-8614
: TELEFAX: 919-541-8689
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 28958 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: US-08-457-335A-6

```

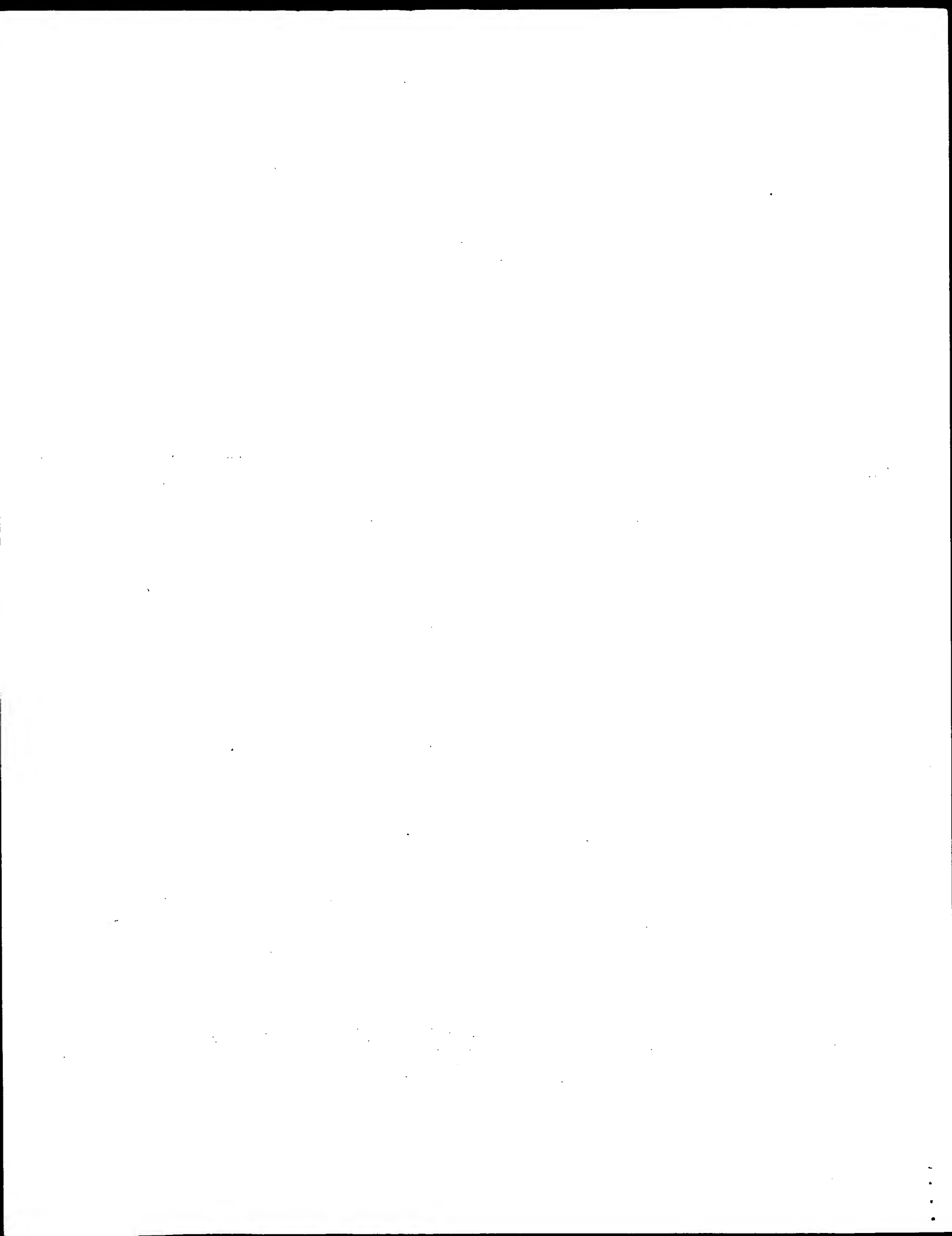
Query Match	4.8%	Score 56;	DB 1;	Length 28958;
Best Local Similarity	48.2%;	Prod. No. 0.0056;		
Matches 158;	Conservative	0;	Mismatches 170;	Indels 0; Gaps 0;
QY	356	caccgcgagacgtgctgtgggcataagccgacctgcgttcctgcagactacgtcgagcc	415	
Db	16356	CGCCCTCTTTGCCGTCATGGTCTCCCTGGCGCCCTCTGGCGGTCTGCTGGGTCTGAGCC	16415	
QY	416	gctcggcgccatactccacgcctaccgcgagattcgaggcgagcgcgcgagctcgagct	475	
Db	16416	CGCCGCGGTGCTGGGCCACACCGAGGCGAGATAGCGCGCCCTTTCGTGCGAGGCGTCT	16475	
QY	476	cgctccggggcgccccatcgcgcgcgcgaccaccaccgcgactccatgtcgccagc	535	
Db	16476	CTTCCCTCGAGACGCGGGCGCCATCGCGCGCCCTGCGCAGCAAGCGTCAACCCGTGCC	16535	
QY	536	gcgatgtctcaagtcctccgcggccagtcctcggagcgcccatgctaccgcgaccacaccca	595	


```

Db 16476 CTCCCTCGAGGACGGCGCGCATCGCGCGCTGGCGAGCAAGCGTCACACCGTCGCC 16535
QY 536 ggcatgtctcaagtccgcgggccaagtctccggagccgccatgtctaccgcaccaccacca 595
Db 16536 GGCAACGGGCAATGGCGCGCGCTCGAGCTCGGGCGCTCCGACCTCCAGACCTACCTCGCTCC 16595
QY 596 ccaccacgacatgcagatgcacgcgcgcctgtacgggggaaacggccgtgccccccgcggc 655
Db 16596 CTGGGGCGACAGGCTCTCCATCGCGCGCGCTCAACAGCCCCAGGGCCACGCTCGTATCCGG 16655
QY 656 cgggctctctaccaccagcggttcttc 683
Db 16656 CGAGCCCGCGCGCTCGACGGCGTGATC 16683

```

Search completed: October 25, 2001, 12:01:59
 Job time: 4344 sec



OM of: US-09-435-054-2 to: Issued Patents_NA.* out_format : pfs

Date: Oct 25, 2001 2:25 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame_b2n.model -DEV=xlh
-Q/cgn2_1/USPTO_Spool/US09435054/runat_25102001_085523_6989/app_query.fasta_1.338
-DB=Issued Patents_NA -QPMT=fastap -SUFFIX=p2n.rni
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-XGAPEXT=0.000 -XGAPOP=4.500 -XGAPEXT=0.050 -XGAPOP=10.000
-YGAPEXT=0.500 -YGAPOP=6.000 -YGAPEXT=7.000 -YGAPOP=10.000
-YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45 -DICALIGN=200
-THR_SCORE=pcr -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=200000000
-USER=US09435054.CGN1_1.77 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-435-054-2

Query length: 278

Database: Issued Patents_NA.*

Database sequences: 324599

Database length: 9465562

Search time (sec): 38.160000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-1 +	494.00	676.81	4.4e-30	627	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-3 +	494.00	660.44	3.6e-29	3395	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-4 +	494.00	652.68	9.7e-29	7560	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-319-989-3 +	295.50	400.48	1.1e-14	835	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-758-662-4 +	167.50	228.50	4.1e-05	530	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-845-998-7 +	167.50	207.72	0.0006	4524	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-206-537-7 +	167.50	207.72	0.0006	4524	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-553-110-2 +	147.00	188.12	0.0073	1853	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-041-886-22 +	147.00	180.00	0.0207	4279	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-041-886-10 +	136.50	166.90	0.1110	3715	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-073-384-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-254-359A-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-483-043-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-481-238-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-471-066B-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-757-653-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-599-491-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-756-386-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-823-516-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-682-853A-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-759-038-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-758-314-2 +	134.50	168.00	0.0964	2496	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-781-802-7 +	134.00	165.07	0.1405	3147	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-694-078-7 +	134.00	165.07	0.1405	3147	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-058-260-7 +	134.00	165.07	0.1405	3147	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-1 +	132.50	166.41	0.1183	2214	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-2 +	132.50	162.45	0.1966	3331	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-864-038A-4 +	132.50	162.45	0.1966	3331	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-581-148C-30 +	129.00	151.38	0.8126	6343	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-30 +	128.50	159.21	0.2977	2634	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-30 +	128.50	159.21	0.2977	2634	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-911-853-29 +	128.50	140.80	3.16	17612	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-479-409-29 +	128.50	140.80	3.16	17612	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-083-351-3 +	125.50	159.56	0.2848	1659	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-352-3 +	125.50	159.56	0.2848	1659	1
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-083-351-1 +	125.50	151.16	0.8362	3946	1
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-083-352-1 +	125.50	151.16	0.8362	3946	1
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-804-227C-1 -	125.50	127.95	16.42	43280	1

/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-928-213B-7 - 124.50 153.53 0.6172 2681
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-690-473-1 + 124.50 149.05 1.10 4257
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-259-821A-1 + 124.50 149.05 1.10 4257
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-843-659-1 + 124.50 149.05 1.10 4257
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-458-568A-11 - 124.50 139.00 3.98 1200

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-103-478-1

seq_documentation_block:
; Sequence 1, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamar
; APPLICANT: Ohno, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,554
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 627 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..627
; OTHER INFORMATION: /product= "LECI"

US-09-103-478-1

alignment_scores:
Quality: 494.00 Length: 249
Ratio: 3.146 Gaps: 9
Percent Similarity: 63.052 Percent Identity: 46.988

alignment_block:
US-09-435-054-2 x US-09-103-478-1 ..

Align seq 1/1 to: US-09-103-478-1 from: 1 to: 627

16 SerAlaAlaGlyClyAlaAsnGlyGlyAlaAlaGlnHisAlaA1 32

```

;
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,478
; FILING DATE: 24-JUN-1998
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/026,221
; FILING DATE: 19-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/804,534
; FILING DATE: 21-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Einhorn, Gregory P.
; REGISTRATION NUMBER: 38,440
; REFERENCE/DOCKET NUMBER: 023070-077611US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 3:
; LENGTH: 3395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-103-478-3

alignment_scores:
    Quality: 494.00      Length: 249
    Ratio: 3.146        Gaps: 9
    Percent Similarity: 63.052    Percent Identity: 46.988

alignment_block:
    US-09-435-054-2 x US-09-103-478-3 ..
    Align seg 1/1 to: US-09-103-478-3 from: 1 to: 3395

16 SerAlaAlaGlyGlyAlaAsnGlyGlyAlaAlaGlnGlnHisAlaAl 32
200 GCGGCGCGCGGTGACAAACAATGGTATCGTGGTCCAGCAGCAACACC 2069
32 aproAlaIleArgGluGlnAspArgLeuMetProIleAlaAsnValIleA 49
2070 ATGTGCTGGCTCGTGAGCAAGCAACATACATGCCAATCGCAACGTCATA 2119
49 rgIleMetArgValLeuProAlaHisAlaLysIleSerAspAla 65
2120 GAATCATGCTAAACCTTACCGTCTCACGCCAAATCTCTGACGACGCC 2169
66 LysGluThrIleGlnGluCysValSerGluTyrIleSerPheIleThrG 82
2170 AAAGAAACGATTCAAGAAATGTGTCTCCGAGTACATCAGCTTCGTCGCCG 2219
82 yGluAlaAsnGluArgCysGlnArgGluGlnArgLysThrIleThrAlaG 99
2220 TGAAGCCAAACGAGCGTTGCCAACGTTGAGCAACGTAAGACCACTAAGCTG 2269
99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
2270 AAGATATCCTTTGGGCTATGAGCAAGCTTGGGTTGCATTAACCTACCTGGAC 2319
116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132

;
; seq_name: /cqn2.6/plotdata/2/1na/6B_COMB.seq:US-09-103-478-3
; seq_documentation_block:
; Sequence 3, Application US/09103478
; Patent No. 6235975
; GENERAL INFORMATION:
; APPLICANT: Harada, John
; APPLICANT: Lotan, Tamir
; APPLICANT: Ohto, Masa-aki
; APPLICANT: Goldberg, Robert B.
; APPLICANT: Fischer, Robert L.
; TITLE OF INVENTION: LEAFY COTYLEDON1 Genes and Their Uses
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:

```

```

alignment_scores:
    Quality: 494.00      Length: 249
    Ratio: 3.146         Gaps: 9
    Percent Similarity: 63.052  Percent Identity: 46.988

alignment_block:
US-09-435-054-2 x US-09-103-478-4 ..
Align seg 1/1 to: US-09-103-478-4 From: 1 to: 7560

16 SerAlaAlaGlyGlyAlaAsnAsnGlyGlyAlaAlaGlnGlnHisAlaAl 32
   ::::||||| ||||| ||||| ||||| ||||| ||||| |||||
4449 GCCGGCGCGGTGACAAAGACAACTGGTATCGTGGTCCAGCAACCAACC 4498
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
32 aProAlaIleArgGluInAspArgLeuMetProIleAlaAsnValIleA 49
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4499 ATGTGTGGCTGTCGAGCAAGACCAATACATACCAATCGCAACGTCATAA 4548
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 rGleMetArgArgValLeuProAlaHisAlaLysIleSerAspAla 65
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4549 GAATCATCGGTAAACACCTTACCGTTCACGCCAAATCTCTCAGCAGCCG 4598
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 LysGluThrIleGlnGlnCysValSerGluTyrIleSerPheIleThrG 82
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4599 AAAGAAACGATTCGAAGATGTCTCCAGTACATCAGCTCTCGACCGG 4648
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 yGluAlaAsnGlnuArgCysGlnArgGlnGlnArgLysThrIleThrAlaG 99
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4649 TGAAGCCACACGAGCGTTCGCCAACGTGAGCAACGTAAAGACCAATACGTG 4698
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 luAspValLeuTrpAlaMetSerArgLeuGlyPheAspAspTyrValGlu 115
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4699 AAGATATCTTTGGGCTATGACAAAGCTTGGTTCGATACTACGTTGAC 4748
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
116 ProLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaAr 132
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4749 CCCCTCACCGTGTTCATTAAACCGGTACCGTACGATGAGATAGACCGCAT...CG 4795
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 gGlyValcIleuValProGlyAlaAlaProSerArgGlyGlyAspHisH 149
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4796 TGGTTCGCACTT.....AGAGT..... 4814
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
149 isProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4815 .....GAGCCACCGTCTGTGACAAACCC..... 4838
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
166 SerGlyAlaAlaMetLeuProHisHisHisHisAspMetGlnMe 182
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4838 ..... 4838
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 tHisAlaAlaMetTyrGlyThrAlaValProProAlaGlyProp 199
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4839 .....TATGAGGAAATGCTATT..... 4856
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
199 roHisHisGlyPheLeuMetProHisProGlnGlySerSerHisTyr 215
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
4857 .....GGGTTT.....CAGGCCCATCATCATGC 4880
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
216 Leu.....ProTyrAlaTyrGluProThrTyrGlyGlyGluHI 228
   :::: ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

422 TAGCTCATTTCTTTTGTGACGTAGCGAGCGCCAGCATCGATCGCGCTGCTG 471
91 luGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerArg 107
472 ACAAGAAGAACGATTAACGGGGAAGACATTCATCATATCATTCGACGCC 521
108 LeuGlyPheAspAspTyrValGluProLeuGlyAlaTyrLeuHisArgTy 124
522 TTAGATTGGAAGAACTATGACAGAGGTGTGAAATCTACTTGGCTAATA 571
124 rArgGlu 126
572 CAGCAA 578

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-758-662-4
seq_documentation_block:
; Sequence 4, Application US/08758662
; Patent No. 6114150
; GENERAL INFORMATION:
; APPLICANT: Weissman, Sherman
; APPLICANT: Baskaran, Namadev
; TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 701 Fifth Avenue, 6300 Columbia Center
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,662
; FILING DATE: 29-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6114150tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 390036.402C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-758-662-4

alignment_scores:
Quality: 167.50 Length: 169
Ratio: 1.971 Gaps: 11
Percent Similarity: 50.296 Percent Identity: 35.503
alignment_block:
US-09-435-054-2 x US-08-758-662-4 ..
Align seg 1/1 to: US-08-758-662-4 from: 1 to: 530
125 ArgGluPheGluGlyAspAlaArgGlyValGly.....Le 136
67 CGAGCGCTGCTGGCGCGCGGAGCGCTGGCGCGCGTGACATCGCGG 116
136 uValProGlyAlaAlaProSer...ArgGlyGlyAspHisHisProHis 152
117 TGTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 166

4881 CTACCTCCCTCGGGTCTTATGGTTAT.....GGTATGTT 4915
228 sAlaMetAlaAlaTyrTyGlyGlyAlaAlaTyrAlaProGlyAsnGlyG 245
4916 GGACCAATCCATGTTATGGGAGGTGGTGGTACTACCAA...AACGGGT 4962
245 lySerGlyAspGlySerGlySerGlyGlyGlyGlyGlySerAlaSer 260
4963 CGTCGGGTCAAGATGAATCAAGTGTGGTGGTGGTCTTCGCTTCC 5009

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-319-989-3
seq_documentation_block:
; Sequence 3, Application US/09319899
; Patent No. 6190914
; GENERAL INFORMATION:
; APPLICANT: Grivell, Leslie A.
; APPLICANT: Teixeira De Mattos, Maarten J.
; APPLICANT: Blom, Jolanda
; TITLE OF INVENTION: METHODS FOR MODULATING METABOLIC PATHWAYS OF
; TITLE OF INVENTION: MICRO-ORGANISMS AND MICRO-ORGANISMS OBTAINABLE BY SAID
; TITLE OF INVENTION: METHODS
; FILE REFERENCE: 24615-20123.00
; CURRENT APPLICATION NUMBER: US/09/319,989
; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: PCT/NL97/00688
; EARLIER FILING DATE: 1997-12-12
; EARLIER APPLICATION NUMBER: EPO 96203520
; EARLIER FILING DATE: 1996-12-12
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 835
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA construct
; OTHER INFORMATION: encoding HAP3
; NAME/KEY: CDS
; LOCATION: (201)..(632)
US-09-319-989-3

alignment_scores:
Quality: 295.50 Length: 119
Ratio: 3.283 Gaps: 2
Percent Similarity: 75.630 Percent Identity: 50.420
alignment_block:
US-09-435-054-2 x US-09-319-989-3 ..
Align seg 1/1 to: US-09-319-989-3 from: 1 to: 835
8 ProAlaAlaGlyAlaGluAsnGlySerAlaAlaGlyGlyAlaAsnAsnG1 24
237 CCAGAGGATACCTCAGGAGACGGT.....GGAAACGCTAGCTCCAG 277
24 yGlyAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArgL 41
278 CGGCAGTTTGCAGCAA.....ATTTCACGCTAAGAGACGAGACAGAT 321
41 euMetProIleAlaAsnValIleArgMetArgValLeuProAla 57
322 GGCTACCATCAACATGTACCGGACTCATGAAGATACTCTCCACCG 371
58 HisAlaTyrIleSerAspAlaAlaGlyGluThrIleGlnGluCysValSe 74
372 AGTGTAAAGGTATCGAAAGATCGCAAGAGTGCATGACGAGGTGTGTCAG 421
74 rCluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArgG 91

```

152 erMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGlyAla 168
    : : : : : : : : : : : : : : : : : : : : : : : :
167 ATGAACAGCGGTCCGTGGACACTCCACTTCCACGCT...GCCTCTGGCGCA 213
    : : : : : : : : : : : : : : : : : : : : : : : :
169 AlaMetLeupro.HisHisHisHisHisHisAspMetGlnMetHisAla 185
    : : : : : : : : : : : : : : : : : : : : : : : :
214 .....CCACCACACACCACCACCAC.....CACCAGG 242
    : : : : : : : : : : : : : : : : : : : : : : : :
185 laMetTyrGlyGlyThrAlaValProProProAlaGlyProHisHis 201
    : : : : : : : : : : : : : : : : : : : : : : : :
243 CGCTC.....GAA 250
    : : : : : : : : : : : : : : : : : : : : : : : :
202 GlyGlyPheLeuMetProHisProGlnGlySerSerHisTyrLeuProTyr 218
    : : : : : : : : : : : : : : : : : : : : : : : :
251 CCGCGGATCTCTGGACCAATCTCTCGCGTCG.....CT 288
    : : : : : : : : : : : : : : : : : : : : : : : :
218 rAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAlaAlaTyrTyrG 235
    : : : : : : : : : : : : : : : : : : : : : : : :
289 CGCGCTCATGGCGCGCGCGCGCGCGCGCG...GCGCGCGCGCGCG 335
    : : : : : : : : : : : : : : : : : : : : : : : :
235 lyGlyAlaAlaTyrAlaProGlyAsn.....243
    : : : : : : : : : : : : : : : : : : : : : : : :
336 GCGGCGCCACACAGCGCCCGGGGGCGGTGGCGCCCGCGCGCGCGCG 385
    : : : : : : : : : : : : : : : : : : : : : : : :
244 .....GlyGlySerGlyAspGlySerGlyGlyGlyGlyGlyGly 258
    : : : : : : : : : : : : : : : : : : : : : : : :
386 GGC CGCGCGCGCGCGCGCGCGCGCGGAGCGCGTGGCGGCCCGCGGGG 435
    : : : : : : : : : : : : : : : : : : : : : : : :
258 rAlaSerHisThrProGlnGlySer.....GlyGlyLeuGluHisProH 273
    : : : : : : : : : : : : : : : : : : : : : : : :
436 CGCGCGCGCGCGCGCGCGCGCGCGGCTCTCTGGCGCGCTCCGCGCACCTC 485
    : : : : : : : : : : : : : : : : : : : : : : : :
273 isPro 274
    : : : : : : : : : : : : : : : : : : : : : : : :
486 ACCCG 490
    : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-845-998-7

```

seq_documentation_block:
; Sequence 7, Application US/08845998
; Patent No. 5879892
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coulic, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,998
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441

```

```
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 4524 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHEetical: NO
;   ANTI-SENSE: NO
;   FEATURE:
;     NAME/KEY: CDS
;     LOCATION: 174..1433
US-08-845-998-7

alignment_scores:
    Quality: 167.50      Length: 169
    Ratio: 1.971        Gaps: 11
    Percent Similarity: 50.296    Percent Identity: 35.503

alignment_block:
US-09-435-054-2 x US-08-845-998-7 ..

Align seg 1/1 to: US-08-845-998-7 from: 1 to: 4524

125 ArgGluPheGluGlyAspAlaArgGlyValGly.....Le 136
||||| ||| ::||| ||| |||
326 CGAGCAGCTGCTGGCGGGCCGAGCGGTGGCGGCCGCGGCATCGCGG 375

136 uValProGlyAlaAlaProSer...ArgGlyGlyAspHisProHis 152
||||||| |||| |:::||| |||
376 TGTCCCAGGCCAAGAGCATCTTCAAGCGGACGCCACGTACCACAGC 425

152 erMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGlyAla 168
||| ::||| ::||| ::||| ::||| ::||| ::|||
426 ATGACAGCGTGCGGTGCAGCTCCATTCACGGT...GCCCTTGGCGCA 472

169 AlaMetLeuPro.HisHisHisHisHisAspMetGlnMetHisAlaA 185
||| ||||| ||||| ||||| |||||
473 .....CCAGCACCAACCACCACCAC.....CACCAGG 501

185 lametTyrgLyGlyThrAlaValProProAlaGlyProProHisHis 201
|||||
502 CGCTC.....GAA 509

202 glyGlyPheLeuMetProHisProGlnGlySerSerHisTyrrProTy 218
||| ||||| ||| ::||| |||
510 CCCGGCATCTGCTGGACCATCTCTCTCGCGCTCG.....CT 547

218 rAlaTyrgLuProThrTyrgLyGlyGlulHisAlaMetAlaTyrrTyg 235
||| ||||| ||||| ||||| |||||
548 CGGGCTCATGGCGCGCGCGGGCGGGCGCG...GCGGCGCGCGCG 594

235 lyGlyAlaAlaTyrrAlaProGlyAsn..... 243
||||||| ::||| |||||
595 GCGGCGCCACGACGCCCCGGGGGGGGTGGCGGCCCGGGCGGGCGGCG 644

244 .....GlyGlySerGlyAspGlySerGlyGlyGlyGlyGlyGly 258
||||| ::||| ||||| ::||| ||||| |||||
645 GGCCCGGGCGGGCGCGCCCGGGGAGGCGCGGTGGCGGCCCGGGGG 694

258 rAlaSerHisThrProGlnGlySer.....GlyGlyLeuGluHisProH 273
::::: ||| ||||| ::||| |||||
695 CGGGCGGGCGGGCGGGCGGGCGGGTCTCTGGCGGGCTCCGCGCACCTC 744

273 lsPro 274
|||||
745 ACCCG 749
```

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-09-206-537-7

seq_documentation_block:

```
seq_documentation.DUCK:
; Sequence 7, Application US/09206537
```

```

; Patent No. 6130052
; GENERAL INFORMATION:
; APPLICANT: Van Baren, Nicolas
; APPLICANT: Coullie, Pierre G.
; APPLICANT: De Smet, Charles
; APPLICANT: Lucas, Sophie
; APPLICANT: Boon, Thierry
; TITLE OF INVENTION: LEUKEMIA ASSOCIATED GENES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
; STREET: 600 Atlantic Avenue
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/206,537
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,998
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Van Amsterdam, John R.
; REGISTRATION NUMBER: 40,212
; REFERENCE/DOCKET NUMBER: L0461/7008
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)720-3500
; TELEFAX: (617)720-2441
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 174..1433
; US-09-206-537-7

alignment_scores:
    Quality: 167.50      Length: 169
    Ratio: 1.971        Gaps: 11
    Percent Similarity: 50.296      Percent Identity: 35.503

alignment_block:
US-09-435-054-2 x US-09-206-537-7 ..
Align seg 1/1 to: US-09-206-537-7 from: 1 to: 4524

125 ArgGluPheGluGlyAspAlaArgGlyValGly.....Le 136
136 CGAGACGCTGCTGGCGGCGCGGCGGTGGCGCGGTGGACATCGCGG 375
136 uValProGlyAlaAlaProSer...ArgGlyGlyAspHisProHis 152
152 TGTCCAGGCAAGACGATCTCTTCAAGCGCGGACGCCAGTACACG 425
152 erMetSerProAlaAlaMetLeuLysSerArgGlyProValSerGlyAla 168
168 ATGAACAGCGTGGCGTGCAGCTCACTTCCACGCT...GCCTCTGGCGCA 472
169 AlaMetLeuPro.HisHisHisHisHisAspMetGlnMetHisAla 185

```

```

473 .....CCACCACACACACACACAC.....CACCAGG 501
185 laMetTyrGlyGlyThrAlaValProProAlaGlyProProHisHis 201
502 CGCTC.....GAA 509
202 GlyGlyPheLeuMetProHisProGlnGlySerSerHisTyrLeuProTyr 218
510 CCAGCGCATCTGTGGACACATCTCTCGCGTGC.....CT 547
218 rAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAlaAlaTyrTyrG 235
548 CGCGCTCATGCGCGCGCGCGCGCGCGCGGCGG...GCAGCGCGCGCGG 594
235 lyGlyAlaAlaTyrAlaProGlyAsn..... 243
595 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 644
244 .....GlyGlySerGlyAspGlySerGlySerGlyGlyGlyGly 258
645 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 694
258 rAlaSerHisThrProGlnGlySer.....GlyGlyLeuGluHisProH 273
695 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 744
273 lsPro 274
745 ACCCG 749

seq_name: /cgn2_5/ptodata/2/ina/5A_COMB.seq:US-08-553-110-2

seq_documentation_block:
; Sequence 2, Application US/08553110
; Patent No. 5723301
; GENERAL INFORMATION:
; APPLICANT: Burke, James R.
; APPLICANT: Vance, Jeffrey M.
; APPLICANT: Enghild, Jan
; APPLICANT: Strittmatter, Warren J.
; TITLE OF INVENTION: Therapeutics for Diseases Associated
; TITLE OF INVENTION: with CAG Trinucleotide Repeat Expansion
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley
; STREET: P.O. Drawer 34009
; CITY: Charlotte
; STATE: NC 27233
; COUNTRY: United States of America
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553,110
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-117
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-420-2200
; TELEFAX: 919-881-3175
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1853 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```


; SOFTWARE: PatentIn Release #1.0, Version #1.25

alignment_scores:	
Quality:	136.50
Ratio:	0.968
Percent Similarity:	45.192
Percent Identity:	23.718
Length:	312
Gaps:	15

```

alignment_block:
US-09-435-054-2 x US-09-041-886-10
Align seg 1/1 to: US-09-041-886-10 from: 1 to: 3715

8  ProAlaAlaGlyAlaGluAsnGlySerAlaAlaGlyAlaAsnAsnG1 24
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1257 CCTGGGTGTGGAGCGTGTGAGCATCTGAGTCCAGGGGAACAGCTTCGGG 1306
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
24  yGlyAlaAlaGlnGlnHisAlaAlaProAlaIleArgGluGlnAspArg. 40
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 GGGATTG.....CATGTACGCCCCACTTTTGGGAGTTCCACCCGCT 1347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
41  .....LeuMetProIleAlaAsnValIleArgIleMetArg 52
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1348 GTCCGTCACCTCTTGTGCCCATCTGGCCGAATGCAMAGGTTCTCTGCT 1397
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
53  Arg.....ValLeuProAl 57
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1398 AGACGACAGCGCAGGCAAGCAGCTGAAGATATCTGTGAGTATTCCTCTT 1447
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
57  aHisAlaLysIleSerAspAlaLysGluThrIleGlnGluCysVals 74
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1448 TCAAGGGAGGTACACCAAGGGCTAGAAAGCGCAGAGCCTAGGCTGCTCT 1497
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
74  erGluTyrIleSerPheIleThrGlyGluAlaAsnGluArgCysGlnArg 90
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1498 GG.....CAGCGCTGCAGCAGG 1514
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
91  GluGlnArgLysThrIleThrAlaGluAspValLeuTrpAlaMetSerAr 107
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1515 GAGCTCGGGACACTTGAAGTGC..... 1538
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
107  gLeuGlyPheAspTyrValGluProLeuGlyAlaTyr.LeuHisArg 123
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1539 .....GTCTACCTGTCTCTC 1554
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
124  TyrArgGluPheGluGlyAspAlaArgGlyValGlyLeuValProGlyAl 140
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1555 TACAAGTCC.....GGAGCACTGGACGAGGC 1580
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
140  aAlaProSerArgGlyAspHisHis..... 149
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1581 AGCTGCTACAGTACGAGTGGGACTACTACAACTTCCACTGGCTCTGGCG 1630
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
150  .....ProHisSerMetSerProAlaAlaMet 158
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1631 GACCGCGCCCTCCGCGCCCTCCCAT.....CCCCAGCTCGC 1671
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
159  LeuLysSerArgGlyProVal.....SerGlyAlaAlaMetLeuPr 172
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1672 ATCAAGCTGAGAACCGCTGGACTACGCGAGCGCTGGCGGCTCGGC 1721
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
172  oHisHisHisHisHisHisAspMetGlnMetHisAlaAlaMetTyrGlyC 189
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1722 GGCAGCTGCGCTATGGGACCTG.....GGAGCCTGCATGGCG 1762
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
189  lyThrAlaValProProAlaGlyProProHisHisGlyGlyPheLeu 205
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1763 CGGTGCGAGCGGACCGGTCTGGGTACCCCTCAGCGCGCT..... 1806
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
206  MetProHisProGlnGlySerHisTyrLeuProTyrAla..... 219
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1807 .....TCTCATCTTGCACACTCTCTTCACAGCCGAGAGG 1844
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
220  .....TyrGluProThrTyrGlyGlyGluHisAlaMetAlaTyrT 234
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1845 CCAGTTGTATGGACCGTGTGTGTGTGG..... 1875
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
234  yrGlyGlyAlaAlaTyrAlaProGlyAsnGlyGlySerGlyAspGlySer 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1876 ..GGTGTGTGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1923
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
251  GlySerGlyGlyGlyGlySerAla..... 259
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1924 GCGCGCGCGCGCGCGCGCGAGAGCTAGTACCCCTACGGCTA 1973
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
260  .SerHisThrProGlnGlySerGlyGlyLeuGlu 270
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1974 CACTCGCGCCCTCAGGGGTGGCGGCGCAGGAA 2007
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/ptodata/2/Ina/5A_COMB.seq:US-08-073-384C-2

seq_documentation_block:
; Sequence 2, Application US/08073384C
; Patent No. 5541311
; GENERAL INFORMATION:
; APPLICANT: Dahlberg, James E.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/073,384C
; FILING DATE: 04-JUN-1993
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-00613
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/705-8410
; TELEFAX: 415/397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-073-384C-2

alignment_scores:
Quality: 134.50 Length: 317
Ratio: 0.934 Gaps: 17
Percent Similarity: 45.426 Percent Identity: 26.498

alignment_block:
US-09-435-054-2 x US-08-073-384C-2
Align seg 1/1 to: US-08-073-384C-2 from: 1 to: 2496

19  GlyGlyAlaAsnGlyGlyAlaAlaGln.GlnHisAlaAlaProAlaI 35
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1035 GGGGTGCGGGGAATCCTGCC.....CAAGGACCTGGCGGTTTGGGCC 1078
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
35  leArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgIleMet 51
   |||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```
1079 TCGGGAGGGCTGACCTCTTCCAGAGGAGCCCATGCTCCTG... 1125
52 ArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGluTh 68
1126 .....GCTACTTCTGACCCCTCCACACAC 1154
68 rIleGlnGluCysValSerGluTyrIleSerPheIleThrGlyAla 85
1155 CCCTGAGGGGTGCGCGGTACGGGGGAGTGCAGCAGGATGCGG 1204
85 snGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAspVal 101
1205 GGGAGAGGGCCCTCTGCGCGAGCGCTCTCCAGACCCCTAAGAGCGC 1254
102 LeuTrpAlaMetSerArgLeu.....GlyPheAspAspTyrValGluPr 116
1255 CTTAAGGGAGAAACCCCTGCTTGTGCTTTACAGCAGGTTGAGAGCC 1304
116 oLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg 133
1305 GCTTTCGGGTGTGGCCCGATGAGGC.....CAGGGGGTCCGGCTG 1350
133 lyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHis 149
1351 GAGTGGCTTACCTCCAGGCCCTCTCCCTGGAGGTGAGCGGAGGTGG 1400
150 ProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal 165
1401 CCA.....GCTGAGGAGGAGGTCTTCCGCCGTGGCGCCACCCCTCA 1444
166 .....SerGlyAlaAlaMetLeu..... 171
1445 ACCTCAACTCCCGGACCACTGGAGGGGTCTTTTACGAGCTGGGC 1494
172 .....ProHisHisHisHisHisAspMetGln...MethHisAlaAla 185
1495 CTGCTGCCATCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1544
186 MetTyrGlyGlyThrAlaValProProAlaGlyProProHisHisG 202
1545 TGGCGTGTGGA.....GCCCTGGCAGAGGCCACCCATCGTGG 1585
202 yGlyPheLeuMetProHisPro.....GlnGlySerS 213
1586 A.....CCGATCTTCACTACGCGGAGCTCACCAGCTCAAG 1623
213 erHisTyr..... 215
1624 AACACCTACATAGACCCCTGCCCGCTGTCCTCCACCCAGACGGCG 1673
216 LeuProTyrAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAla 231
1674 GCTCCACACCCGCTTCAACACAG...GCCACCGCCAGCGGAGGCT 1720
232 .....AlaTyrTyrGlyGlyA 237
1721 CCAGCTCCGACCCCAACCTGCAGAACATCCCGTGGCACCCTCTGGC 1770
237 laAlaTyrAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
1771 CAGCGCATCCGCGAGCCCTTCGTGCGCAGAGAGGCTGGTCTGTGTGT 1820
250 .....SerGlySerGlyGlyGlyGlyGlySe 258
1821 CTGGACTACACCATGACATGACCTTGGGTCTGCCACCTCTCCGGGG 1870
258 rAlaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
1871 ACGAACACCTGATCGGCTTTCAGGAGGAGGAGGACATCCACACCCA 1919
seq_name: /cgn2_6/ptdata/2/ina/5A_COMB.seq.us-08-254-359A-2
seq_documentation_block:
```

```
; Sequence 2, Application US/08254359A
; Patent No. 5614402
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,359A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/073,384
; FILING DATE: 06-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/986,330
; FILING DATE: 07-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-254-359A-2
```

alignment_scores:

```
Quality: 134.50 Length: 317
Ratio: 0.934 Gaps: 17
Percent Similarity: 45.426 Percent Identity: 26.498
```

alignment_block:

```
US-09-435-054-2 x US-08-254-359A-2 ..
Align seg 1/1 to: US-08-254-359A-2 from: 1 to: 2496
19 GlyGlyAlaAsnAsnGlyGlyAlaAlaGln.GlnHisAlaAlaProAlaI 35
1035 GGGGGTGGGGGATCTGCG.....CAAGGACCTGGCGGTTTGGCCC 1078
35 leArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgIleMet 51
1079 TCGGGAGGGCTGACCTCTTCCAGAGGAGCCCATGCTCCTG... 1125
52 ArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGluTh 68
1126 .....GCCCTACTTCTGACCCCTCCACACAC 1154
68 rIleGlnGluCysValSerGluTyrIleSerPheIleThrGlyGluAla 85
1155 CCCTGAGGGGTGCGCGGTACGGGGGAGTGCAGCAGGATGCGG 1204
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-483-043-2

```

1305 GCTTCCCGGGTGTGGCCGGATGAGGC.....CACGGGGTCCGGCTG 1350
133 lyValGlyLeuValProGlyAlaAProSerArgGlyGlyAspHisHis 149
1351 GACGTGGCCTACCTCCAGGCCCTCTCCCTGGAGGTGGAGGGAGGTGGC 1400
150 ProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal... 165
1401 CCA.....GCTGGAGGAGGAGGTCTTCCCGCTGGCCGCCACCCCTTCA 1444
166 .....SerGlyAlaAlaMetLeu..... 171
1445 ACCTCACTCCCGGACACGCTGGAGCGGTGCTCTTTGACGAGCTGGGC 1494
172 .....ProHisHisHisHisHisHisMetGln...MetHisAlaAla 185
1495 CTGCCTGCCATCGCAAGAGGAGGAGAGGAGGAGGAGGAGGAGGAGG 1544
186 MetTyrGlyGlyThrAlaValProProAlaGlyProHisHisG1 202
1545 TCCCGTGTGGA.....GGCCCTGGAGAGGCCACCCCATCGTGG 1585
202 yGlyPheLeuMetProHisPro.....GlnGlySers 213
1586 A.....CCGCATCTGTCAGTACCGGGAGCTCACCAAGCTCAAG 1623
213 erHisTyr..... 215
1624 AACACCTATAGACCCCTCCCGCCCTGTGTCACCCCAAGACCGGCGG 1673
216 LeuProTyrAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAla.. 231
1674 GCTCCACACCCGCTTCAACACAGAC...GGCCACCGCCAGCGGAGGCTT 1720
232 .....AlaTyrTyrGlyGlyA 237
1721 CGAGCTCCGACCCCAACCTGCAGAACATCCCGCTGGCAGCCCTCTGGC 1770
237 laAlaTyrAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
1771 CAGCGCATCCGCGGAGCCTTCGTGCGGAGGAGGAGGAGGAGGAGGAG 1820
250 .....SerGlySerGlyGlyGlyGlySe 258
1821 CTTGGACTACAGCCAGATTGAGCTTCGGGTCTGCGCCACCTCTCCGGG 1870
258 rAlaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
1871 ACAGAACTGATCCGGGTCTTTCAGGAGGAGGAGGAGGAGGAGGAGGAG 1919

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-481-238-2

seq_documentation_block:
; Sequence 2, Application US/08481238
; Patent No. 5795763
; GENERAL INFORMATION:
; APPLICANT: DAHLBERG, JAMES E.
; APPLICANT: LYAMICHEV, VICTOR I.
; APPLICANT: BROW, MARY ANN D.
; TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
; NUMBER OF INVENTION: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,238
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01798
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2496 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-481-238-2

alignment_scores:
Quality: 134.50 Length: 317
Ratio: 0.934 Gaps: 17
Percent Similarity: 45.426 Percent Identity: 26.496

alignment_block:
US-09-435-054-2 x US-08-481-238-2 ..
Align seg 1/1 to: US-08-481-238-2 from: 1 to: 2496
19 GlyGlyAlaAsnAsnGlyAlaAlaGln.GlnHisAlaAlaProAlaI 35
1035 GGGGGTGGGGGAATCCTGGC.....CAAGGACCTGGCGTTTGGCC 1078
35 leArgGluGlnAspArgLeuMetProIleAlaAsnValIleArgIleMet 51
1079 TGGCGGAGGGCCCTGGACCTCTTCCAGAGGACGACCCCATGCTCTG... 1125
52 ArgArgValLeuProAlaHisAlaLysIleSerAspAlaLysGluTh 68
1126 .....GCTACCTCTCTGGACCCCTCCCAACACAC 1154
68 rIleGlnGluCysValSerGluTyrIleSerPheIleThrGlyGluAla 85
1155 CCCTGAGGGGGTGGCCGGCTTACGGGGGGAGTGGACGAGGATGCGG 1204
85 snGluArgCysGlnArgGluGlnArgLysThrIleThrAlaGluAspVal 101
1205 GGGAGAGGGCCCTCTCGCCGAGCGCTCTTCCAGACCCCTAAAGGAGCG 1254
102 LeuTrpAlaMetSerArgLeu.....GlyPheAspAspTyrValGluPr 116
1255 CTTAAGGGAGAGAAGACGCTCTTGGCTTTAGGAGGAGGTGGAGAGCC 1304
116 oLeuGlyAlaTyrLeuHisArgTyrArgGluPheGluGlyAspAlaArg 133
1305 GCTTCCCGGGGTGTGGCCGGATGAGGC....CACGGGGTCCGGCTG 1350
133 lyValGlyLeuValProGlyAlaAlaProSerArgGlyGlyAspHisHis 149
1351 GACGTGGCCTACCTCCAGGCCCTCTCCCTGGAGGTGGAGGGAGGTGGC 1400
150 ProHisSerMetSerProAlaAlaMetLeuLysSerArgGlyProVal... 165
1401 CCA.....GCTGGAGGAGGAGGTCTTCCCGCTGGCCGCCACCCCTCA 1444
166 .....SerGlyAlaAlaMetLeu..... 171
1445 ACCTCACTCCCGGACACGCTGGAGCGGTGCTCTTTGACGAGCTGGGC 1494

```

172ProHisHisHisHisHisAspMetGln....MetHisAlaAla 185
|||:||||| :||| :|||
1495 CTGCCTGCCATCGGCAAGACGAGAGAGAGGGGAAACGCTCCACCACGGCG 1544
186 MetTyrGlyGlyThrAlaValProProAlaGlyProProHisHisS1 202
:||||| :||| :|||||:|||||:|||||
1545 TGCCGTGCTGGA.....GGCCTGGGAGAGGCCACCCTCATCGTGG 1585
202 yGlyPheLeuMetProHisPro.....GlnGlySers 213
| :||||| :|||
1586 A.....CGGCATCTCTGCAGTACCGGGAGCTCACCAAGCTCAAG 1623
213 erHistyr..... 215
:|||||
1624 AACCTACATAGACCCCTGCCGCCCTGTCTCCACCCCAAGACCGGCG 1673
216 LeuProTyrAlaTyrGluProThrTyrGlyGlyGluHisAlaMetAla.. 231
|||:| :||| :||| :|||||:|||||
1674 GCTCCACACCCGCTTCAACCGAC...GCCACCGCCACGGGCGGCTTT 1720
232AlaTyrTyrGlyGlyA 237
|||:| :|||
1721 CCAGCTCCGACCCCAACCTGCAGAACATCCCGCTGCGCACCCCTCTGGGC 1770
237 laAlaTyrAlaPro.....GlyAsnGlyGlySerGlyAspGly... 249
|||:| :||| :|||||:||||| :|||
1771 CAGCGCATCCCGCGAGCCTTCGTGCGCCGAGGAGGGCTGGGTGCTGGT 1820
250SerGlySerGlyGlyGlyGlyGlyse 258
||||| :|||
1821 CTTGGACTACAGCCAGATTGAGCTTCGGGTCTTGGCCCACTCTCCGGGG 1870
258 rAlaSerHisThrProGlnGlySerGlyGlyLeuGluHisProHisPro 274
: :||| :||||| :|||||
1871 ACGAAGACCTGATCCGGGTCTTTAGGAGAGGGGAGGACATCCACACCCA 1919

